

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 12:00:59 ; Search time 17.6774 Seconds  
(without alignments)  
364.457 Million cell updates/sec

Title: US-09-941-314-2

Perfect score: 1 MAEPWQALQLLAILTLTMA.....VFAPWPEQYKILNKSCSSD 137

Sequence: BLOSUM62

Scoring table: Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_41.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	736	100.0	137	1 CS11_HUMAN	Q9H112 homo sapien
2	422	57.3	139	1 CS11_HUMAN	Q9H112 mus musculus
3	226.5	30.8	142	1 CS11_HUMAN	Q9H112 mus musculus
4	202	27.4	142	1 CS11_HUMAN	Q9H112 mus musculus
5	202	27.4	142	1 CS11_HUMAN	Q9H112 mus musculus
6	186.5	25.3	140	1 CS11_HUMAN	Q9H112 mus musculus
7	184.5	25.1	148	1 CS11_HUMAN	Q9H112 mus musculus
8	181	24.6	146	1 CS11_HUMAN	Q9H112 mus musculus
9	174	23.6	141	1 CS11_HUMAN	Q9H112 mus musculus
10	173.5	23.6	139	1 CS11_HUMAN	Q9H112 mus musculus
11	171.5	23.2	127	1 CS11_HUMAN	Q9H112 mus musculus
12	171	23.2	146	1 CS11_HUMAN	Q9H112 mus musculus
13	170	23.1	146	1 CS11_HUMAN	Q9H112 mus musculus
14	169.5	23.0	116	1 CS11_HUMAN	Q9H112 mus musculus
15	166.5	22.6	165	1 CS11_HUMAN	Q9H112 mus musculus
16	165.5	22.5	148	1 CS11_HUMAN	Q9H112 mus musculus
17	161	21.9	141	1 CS11_HUMAN	Q9H112 mus musculus
18	153.5	20.9	141	1 CS11_HUMAN	Q9H112 mus musculus
19	147.5	20.0	147	1 CS11_HUMAN	Q9H112 mus musculus
20	146.5	19.9	142	1 CS11_HUMAN	Q9H112 mus musculus
21	143	19.4	129	1 CS11_HUMAN	Q9H112 mus musculus
22	136.5	18.5	137	1 CS11_HUMAN	Q9H112 mus musculus
23	135	18.3	149	1 CS11_HUMAN	Q9H112 mus musculus
24	133	18.1	130	1 CS11_HUMAN	Q9H112 mus musculus
25	130	17.7	130	1 CS11_HUMAN	Q9H112 mus musculus
26	129.5	17.6	141	1 CS11_HUMAN	Q9H112 mus musculus
27	127	17.3	111	1 CS11_HUMAN	Q9H112 mus musculus
28	109	14.8	145	1 CS11_HUMAN	Q9H112 mus musculus
29	108	14.7	144	1 CS11_HUMAN	Q9H112 mus musculus
30	98.5	13.4	434	1 CS11_HUMAN	Q9H112 mus musculus
31	98.5	13.4	619	1 CS11_HUMAN	Q9H112 mus musculus
32	96.5	13.1	436	1 CS11_HUMAN	Q9H112 mus musculus
33	96.5	13.1	621	1 CS11_HUMAN	Q9H112 mus musculus

34	95.5	13.0	162	1 CS11_HUMAN	Q9H112 mus musculus
35	93.5	12.7	644	1 CS11_HUMAN	Q9H112 mus musculus
36	87	11.8	639	1 CS11_HUMAN	Q9H112 mus musculus
37	81	11.0	430	1 CS11_HUMAN	Q9H112 mus musculus
38	79	10.7	430	1 CS11_HUMAN	Q9H112 mus musculus
39	79	10.7	661	1 CS11_HUMAN	Q9H112 mus musculus
40	77.5	10.5	504	1 CS11_HUMAN	Q9H112 mus musculus
41	77	10.5	214	1 CS11_HUMAN	Q9H112 mus musculus
42	77	10.5	430	1 CS11_HUMAN	Q9H112 mus musculus
43	75.5	10.3	400	1 CS11_HUMAN	Q9H112 mus musculus
44	73.5	10.0	438	1 CS11_HUMAN	Q9H112 mus musculus
45	73	9.9	861	1 CS11_HUMAN	Q9H112 mus musculus

## ALIGNMENTS

RESULT 1  
ID CS11\_HUMAN STANDARD; PRT; 137 AA.  
AC Q9H112; Q9H113;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Cystatin II precursor.  
GN CystII OR CystII.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RX MEDLINE=21638749; PubMed=11780052;  
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
RA Jones M., Stevrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
RA Bailey J., Barlow K.P., Bates K.N., Beard L.M., Beare D.M.,  
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgman A.M., Brown A.J.,  
RA Buck D., Burrill W.D., Butler A.P., Carter C., Carter N.P.,  
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
RA Coulson A., Coville G.J., Deadman R., Dham P.D., Dunn M.,  
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
RA Graffham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
RA Kay M.P., Kimberley A.M., King A., Knight A., Laird G.R., Lawlor S.,  
RA Leheslahti M.H., Leverhwa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
RA Marsh V.L., Martin S.L., McConachie L.J., McKay K., McMurtry A.A.,  
RA Milne S.A., Mistry D., Moore M.U.P., Mullikin J.C., Nickerson T.,  
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H.,  
RA Rice C.M., Rose M.T., Scott C.B., Sehra H.K., Showkhen R., Sims S.,  
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulten J.E.,  
RA Swann R.M., Symcox N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Williams S.A.,  
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Beck S.,  
RA Williams L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
RA Rogers J.;  
RT "The DNA sequence and comparative analysis of human chromosome 20.";  
RL Nature 414:865-871(2001).  
CC -1- SUBCELLULAR LOCATION: Secreted (Potential).  
CC -1- ALTERNATIVE PRODUCTS:  
CC Name=1;  
CC Name=2;  
CC Name=3;  
CC Name=4;  
CC Name=5;  
CC Name=6;  
CC Name=7;  
CC Name=8;  
CC Name=9;  
CC Name=10;  
CC Name=11;  
CC Name=12;  
CC Name=13;  
CC Name=14;  
CC Name=15;  
CC Name=16;  
CC Name=17;  
CC Name=18;  
CC Name=19;  
CC Name=20;  
CC Name=21;  
CC Name=22;  
CC Name=23;  
CC Name=24;  
CC Name=25;  
CC Name=26;  
CC Name=27;  
CC Name=28;  
CC Name=29;  
CC Name=30;  
CC Name=31;  
CC Name=32;  
CC Name=33;  
CC Name=34;  
CC Name=35;  
CC Name=36;  
CC Name=37;  
CC Name=38;  
CC Name=39;  
CC Name=40;  
CC Name=41;  
CC Name=42;  
CC Name=43;  
CC Name=44;  
CC Name=45;  
CC Name=46;  
CC Name=47;  
CC Name=48;  
CC Name=49;  
CC Name=50;  
CC Name=51;  
CC Name=52;  
CC Name=53;  
CC Name=54;  
CC Name=55;  
CC Name=56;  
CC Name=57;  
CC Name=58;  
CC Name=59;  
CC Name=60;  
CC Name=61;  
CC Name=62;  
CC Name=63;  
CC Name=64;  
CC Name=65;  
CC Name=66;  
CC Name=67;  
CC Name=68;  
CC Name=69;  
CC Name=70;  
CC Name=71;  
CC Name=72;  
CC Name=73;  
CC Name=74;  
CC Name=75;  
CC Name=76;  
CC Name=77;  
CC Name=78;  
CC Name=79;  
CC Name=80;  
CC Name=81;  
CC Name=82;  
CC Name=83;  
CC Name=84;  
CC Name=85;  
CC Name=86;  
CC Name=87;  
CC Name=88;  
CC Name=89;  
CC Name=90;  
CC Name=91;  
CC Name=92;  
CC Name=93;  
CC Name=94;  
CC Name=95;  
CC Name=96;  
CC Name=97;  
CC Name=98;  
CC Name=99;  
CC Name=100;  
CC Name=101;  
CC Name=102;  
CC Name=103;  
CC Name=104;  
CC Name=105;  
CC Name=106;  
CC Name=107;  
CC Name=108;  
CC Name=109;  
CC Name=110;  
CC Name=111;  
CC Name=112;  
CC Name=113;  
CC Name=114;  
CC Name=115;  
CC Name=116;  
CC Name=117;  
CC Name=118;  
CC Name=119;  
CC Name=120;  
CC Name=121;  
CC Name=122;  
CC Name=123;  
CC Name=124;  
CC Name=125;  
CC Name=126;  
CC Name=127;  
CC Name=128;  
CC Name=129;  
CC Name=130;  
CC Name=131;  
CC Name=132;  
CC Name=133;  
CC Name=134;  
CC Name=135;  
CC Name=136;  
CC Name=137;  
CC Name=138;  
CC Name=139;  
CC Name=140;  
CC Name=141;  
CC Name=142;  
CC Name=143;  
CC Name=144;  
CC Name=145;  
CC Name=146;  
CC Name=147;  
CC Name=148;  
CC Name=149;  
CC Name=150;  
CC Name=151;  
CC Name=152;  
CC Name=153;  
CC Name=154;  
CC Name=155;  
CC Name=156;  
CC Name=157;  
CC Name=158;  
CC Name=159;  
CC Name=160;  
CC Name=161;  
CC Name=162;  
CC Name=163;  
CC Name=164;  
CC Name=165;  
CC Name=166;  
CC Name=167;  
CC Name=168;  
CC Name=169;  
CC Name=170;  
CC Name=171;  
CC Name=172;  
CC Name=173;  
CC Name=174;  
CC Name=175;  
CC Name=176;  
CC Name=177;  
CC Name=178;  
CC Name=179;  
CC Name=180;  
CC Name=181;  
CC Name=182;  
CC Name=183;  
CC Name=184;  
CC Name=185;  
CC Name=186;  
CC Name=187;  
CC Name=188;  
CC Name=189;  
CC Name=190;  
CC Name=191;  
CC Name=192;  
CC Name=193;  
CC Name=194;  
CC Name=195;  
CC Name=196;  
CC Name=197;  
CC Name=198;  
CC Name=199;  
CC Name=200;  
CC Name=201;  
CC Name=202;  
CC Name=203;  
CC Name=204;  
CC Name=205;  
CC Name=206;  
CC Name=207;  
CC Name=208;  
CC Name=209;  
CC Name=210;  
CC Name=211;  
CC Name=212;  
CC Name=213;  
CC Name=214;  
CC Name=215;  
CC Name=216;  
CC Name=217;  
CC Name=218;  
CC Name=219;  
CC Name=220;  
CC Name=221;  
CC Name=222;  
CC Name=223;  
CC Name=224;  
CC Name=225;  
CC Name=226;  
CC Name=227;  
CC Name=228;  
CC Name=229;  
CC Name=230;  
CC Name=231;  
CC Name=232;  
CC Name=233;  
CC Name=234;  
CC Name=235;  
CC Name=236;  
CC Name=237;  
CC Name=238;  
CC Name=239;  
CC Name=240;  
CC Name=241;  
CC Name=242;  
CC Name=243;  
CC Name=244;  
CC Name=245;  
CC Name=246;  
CC Name=247;  
CC Name=248;  
CC Name=249;  
CC Name=250;  
CC Name=251;  
CC Name=252;  
CC Name=253;  
CC Name=254;  
CC Name=255;  
CC Name=256;  
CC Name=257;  
CC Name=258;  
CC Name=259;  
CC Name=260;  
CC Name=261;  
CC Name=262;  
CC Name=263;  
CC Name=264;  
CC Name=265;  
CC Name=266;  
CC Name=267;  
CC Name=268;  
CC Name=269;  
CC Name=270;  
CC Name=271;  
CC Name=272;  
CC Name=273;  
CC Name=274;  
CC Name=275;  
CC Name=276;  
CC Name=277;  
CC Name=278;  
CC Name=279;  
CC Name=280;  
CC Name=281;  
CC Name=282;  
CC Name=283;  
CC Name=284;  
CC Name=285;  
CC Name=286;  
CC Name=287;  
CC Name=288;  
CC Name=289;  
CC Name=290;  
CC Name=291;  
CC Name=292;  
CC Name=293;  
CC Name=294;  
CC Name=295;  
CC Name=296;  
CC Name=297;  
CC Name=298;  
CC Name=299;  
CC Name=300;  
CC Name=301;  
CC Name=302;  
CC Name=303;  
CC Name=304;  
CC Name=305;  
CC Name=306;  
CC Name=307;  
CC Name=308;  
CC Name=309;  
CC Name=310;  
CC Name=311;  
CC Name=312;  
CC Name=313;  
CC Name=314;  
CC Name=315;  
CC Name=316;  
CC Name=317;  
CC Name=318;  
CC Name=319;  
CC Name=320;  
CC Name=321;  
CC Name=322;  
CC Name=323;  
CC Name=324;  
CC Name=325;  
CC Name=326;  
CC Name=327;  
CC Name=328;  
CC Name=329;  
CC Name=330;  
CC Name=331;  
CC Name=332;  
CC Name=333;  
CC Name=334;  
CC Name=335;  
CC Name=336;  
CC Name=337;  
CC Name=338;  
CC Name=339;  
CC Name=340;  
CC Name=341;  
CC Name=342;  
CC Name=343;  
CC Name=344;  
CC Name=345;  
CC Name=346;  
CC Name=347;  
CC Name=348;  
CC Name=349;  
CC Name=350;  
CC Name=351;  
CC Name=352;  
CC Name=353;  
CC Name=354;  
CC Name=355;  
CC Name=356;  
CC Name=357;  
CC Name=358;  
CC Name=359;  
CC Name=360;  
CC Name=361;  
CC Name=362;  
CC Name=363;  
CC Name=364;  
CC Name=365;  
CC Name=366;  
CC Name=367;  
CC Name=368;  
CC Name=369;  
CC Name=370;  
CC Name=371;  
CC Name=372;  
CC Name=373;  
CC Name=374;  
CC Name=375;  
CC Name=376;  
CC Name=377;  
CC Name=378;  
CC Name=379;  
CC Name=380;  
CC Name=381;  
CC Name=382;  
CC Name=383;  
CC Name=384;  
CC Name=385;  
CC Name=386;  
CC Name=387;  
CC Name=388;  
CC Name=389;  
CC Name=390;  
CC Name=391;  
CC Name=392;  
CC Name=393;  
CC Name=394;  
CC Name=395;  
CC Name=396;  
CC Name=397;  
CC Name=398;  
CC Name=399;  
CC Name=400;  
CC Name=401;  
CC Name=402;  
CC Name=403;  
CC Name=404;  
CC Name=405;  
CC Name=406;  
CC Name=407;  
CC Name=408;  
CC Name=409;  
CC Name=410;  
CC Name=411;  
CC Name=412;  
CC Name=413;  
CC Name=414;  
CC Name=415;  
CC Name=416;  
CC Name=417;  
CC Name=418;  
CC Name=419;  
CC Name=420;  
CC Name=421;  
CC Name=422;  
CC Name=423;  
CC Name=424;  
CC Name=425;  
CC Name=426;  
CC Name=427;  
CC Name=428;  
CC Name=429;  
CC Name=430;  
CC Name=431;  
CC Name=432;  
CC Name=433;  
CC Name=434;  
CC Name=435;  
CC Name=436;  
CC Name=437;  
CC Name=438;  
CC Name=439;  
CC Name=440;  
CC Name=441;  
CC Name=442;  
CC Name=443;  
CC Name=444;  
CC Name=445;  
CC Name=446;  
CC Name=447;  
CC Name=448;  
CC Name=449;  
CC Name=450;  
CC Name=451;  
CC Name=452;  
CC Name=453;  
CC Name=454;  
CC Name=455;  
CC Name=456;  
CC Name=457;  
CC Name=458;  
CC Name=459;  
CC Name=460;  
CC Name=461;  
CC Name=462;  
CC Name=463;  
CC Name=464;  
CC Name=465;  
CC Name=466;  
CC Name=467;  
CC Name=468;  
CC Name=469;  
CC Name=470;  
CC Name=471;  
CC Name=472;  
CC Name=473;  
CC Name=474;  
CC Name=475;  
CC Name=476;  
CC Name=477;  
CC Name=478;  
CC Name=479;  
CC Name=480;  
CC Name=481;  
CC Name=482;  
CC Name=483;  
CC Name=484;  
CC Name=485;  
CC Name=486;  
CC Name=487;  
CC Name=488;  
CC Name=489;  
CC Name=490;  
CC Name=491;  
CC Name=492;  
CC Name=493;  
CC Name=494;  
CC Name=495;  
CC Name=496;  
CC Name=497;  
CC Name=498;  
CC Name=499;  
CC Name=500;  
CC Name=501;  
CC Name=502;  
CC Name=503;  
CC Name=504;  
CC Name=505;  
CC Name=506;  
CC Name=507;  
CC Name=508;  
CC Name=509;  
CC Name=510;  
CC Name=511;  
CC Name=512;  
CC Name=513;  
CC Name=514;  
CC Name=515;  
CC Name=516;  
CC Name=517;  
CC Name=518;  
CC Name=519;  
CC Name=520;  
CC Name=521;  
CC Name=522;  
CC Name=523;  
CC Name=524;  
CC Name=525;  
CC Name=526;  
CC Name=527;  
CC Name=528;  
CC Name=529;  
CC Name=530;  
CC Name=531;  
CC Name=532;  
CC Name=533;  
CC Name=534;  
CC Name=535;  
CC Name=536;  
CC Name=537;  
CC Name=538;  
CC Name=539;  
CC Name=540;  
CC Name=541;  
CC Name=542;  
CC Name=543;  
CC Name=544;  
CC Name=545;  
CC Name=546;  
CC Name=547;  
CC Name=548;  
CC Name=549;  
CC Name=550;  
CC Name=551;  
CC Name=552;  
CC Name=553;  
CC Name=554;  
CC Name=555;  
CC Name=556;  
CC Name=557;  
CC Name=558;  
CC Name=559;  
CC Name=560;  
CC Name=561;  
CC Name=562;  
CC Name=563;  
CC Name=564;  
CC Name=565;  
CC Name=566;  
CC Name=567;  
CC Name=568;  
CC Name=569;  
CC Name=570;  
CC Name=571;  
CC Name=572;  
CC Name=573;  
CC Name=574;  
CC Name=575;  
CC Name=576;  
CC Name=577;  
CC Name=578;  
CC Name=579;  
CC Name=580;  
CC Name=581;  
CC Name=582;  
CC Name=583;  
CC Name=584;  
CC Name=585;  
CC Name=586;  
CC Name=587;  
CC Name=588;  
CC Name=589;  
CC Name=590;  
CC Name=591;  
CC Name=592;  
CC Name=593;  
CC Name=594;  
CC Name=595;  
CC Name=596;  
CC Name=597;  
CC Name=598;  
CC Name=599;  
CC Name=600;  
CC Name=601;  
CC Name=602;  
CC Name=603;  
CC Name=604;  
CC Name=605;  
CC Name=606;  
CC Name=607;  
CC Name=608;  
CC Name=609;  
CC Name=610;  
CC Name=611;  
CC Name=612;  
CC Name=613;  
CC Name=614;  
CC Name=615;  
CC Name=616;  
CC Name=617;  
CC Name=618;  
CC Name=619;  
CC Name=620;  
CC Name=621;  
CC Name=622;  
CC Name=623;  
CC Name=624;  
CC Name=625;  
CC Name=626;  
CC Name=627;  
CC Name=628;  
CC Name=629;  
CC Name=630;  
CC Name=631;  
CC Name=632;  
CC Name=633;  
CC Name=634;  
CC Name=635;  
CC Name=636;  
CC Name=637;  
CC Name=638;  
CC Name=639;  
CC Name=640;  
CC Name=641;  
CC Name=642;  
CC Name=643;  
CC Name=644;  
CC Name=645;  
CC Name=646;  
CC Name=647;  
CC Name=648;  
CC Name=649;  
CC Name=650;  
CC Name=651;  
CC Name=652;  
CC Name=653;  
CC Name=654;  
CC Name=655;  
CC Name=656;  
CC Name=657;  
CC Name=658;  
CC Name=659;  
CC Name=660;  
CC Name=661;  
CC Name=662;  
CC Name=663;  
CC Name=664;  
CC Name=665;  
CC Name=666;  
CC Name=667;  
CC Name=668;  
CC Name=669;  
CC Name=670;  
CC Name=671;  
CC Name=672;  
CC Name=673;  
CC Name=674;  
CC Name=675;  
CC Name=676;  
CC Name=677;  
CC Name=678;  
CC Name=679;  
CC Name=680;  
CC Name=681;  
CC Name=682;  
CC Name=683;  
CC Name=684;  
CC Name=685;  
CC Name=686;  
CC Name=687;  
CC Name=688;  
CC Name=689;  
CC Name=690;  
CC Name=691;  
CC Name=692;  
CC Name=693;  
CC Name=694;  
CC Name=695;  
CC Name=696;  
CC Name=697;  
CC Name=698;  
CC Name=699;  
CC Name=700;  
CC Name=701;  
CC Name=702;  
CC Name=703;  
CC Name=704;  
CC Name=705;  
CC Name=706;  
CC Name=707;  
CC Name=708;  
CC Name=709;  
CC Name=710;  
CC Name=711;  
CC Name=712;  
CC Name=713;  
CC Name=714;  
CC Name=715;  
CC Name=716;  
CC Name=717;  
CC Name=718;  
CC Name=719;  
CC Name=720;  
CC Name=721;  
CC Name=722;  
CC Name=723;  
CC Name=724;  
CC Name=725;  
CC Name=726;  
CC Name=727;  
CC Name=728;  
CC Name=729;  
CC Name=730;  
CC Name=731;  
CC Name=732;  
CC Name=733;  
CC Name=734;  
CC Name=735;  
CC Name=736;  
CC Name=737;  
CC Name=738;  
CC Name=739;  
CC Name=740;  
CC Name=741;  
CC Name=742;  
CC Name=743;  
CC Name=744;  
CC Name=745;  
CC Name=746;  
CC Name=747;  
CC Name=748;  
CC Name=749;  
CC Name=750;  
CC Name=751;  
CC Name=752;  
CC Name=753;  
CC Name=754;  
CC Name=755;  
CC Name=756;  
CC Name=757;  
CC Name=758;  
CC Name=759;  
CC Name=760;  
CC Name=761;  
CC Name=762;  
CC Name=763;  
CC Name=764;  
CC Name=765;  
CC Name=766;  
CC Name=767;  
CC Name=768;  
CC Name=769;  
CC Name=770;  
CC Name=771;  
CC Name=772;  
CC Name=773;  
CC Name=774;  
CC Name=775;  
CC Name=776;  
CC Name=777;  
CC Name=778;  
CC Name=779;  
CC Name=780;  
CC Name=781;  
CC Name=782;  
CC Name=783;  
CC Name=784;  
CC Name=785;  
CC Name=786;  
CC Name=787;  
CC Name=788;  
CC Name=789;  
CC Name=790;  
CC Name=791;  
CC Name=792;  
CC Name=793;  
CC Name=794;  
CC Name=795;  
CC Name=796;  
CC Name=797;  
CC Name=798;  
CC Name=799;  
CC Name=800;  
CC Name=801;  
CC Name=802;  
CC Name=803;  
CC Name=804;  
CC Name=805;  
CC Name=806;  
CC Name=807;  
CC Name=808;  
CC Name=809;  
CC Name=810;  
CC Name=811;  
CC Name=812;  
CC Name=813;  
CC Name=814;  
CC Name=815;  
CC Name=816;  
CC Name=817;  
CC Name=818;  
CC Name=819;  
CC Name=820;  
CC Name=821;  
CC Name=822;  
CC Name=823;  
CC Name=824;  
CC Name=825;  
CC Name=826;  
CC Name=827;  
CC Name=82

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; AL096677; CAC13170.1; -  
 DR EMBL; AL096677; CAC17423.1; -  
 DR HSSP; P01038; 1A90.  
 DR Genew; HGNC:15959; CST11.  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF00031; Cystatin; 1.  
 DR SMART; SM00043; CY; 1.  
 DR PROSITE; PS00287; CYSTATIN; FALSE NEG.  
 KM Thiol protease inhibitor; Signal; Alternative splicing.  
 FT SIGNAL 1 25  
 FT CHAIN 26 137  
 FT SITE 75 79  
 FT DISULFID 93 101  
 FT DISULFID 114 134  
 FT CARBOHYD 131 131  
 FT VARSPLIC 76 110  
 FT SEQUENCE 137 AA; 16375 MW; C585C8C39A585C3B CRC64;  
 SQ

Query March 100.0%; Score 736; DB 1; Length 137;  
 Best Local Similarity 100.0%; Pred. No. 1,7e-65;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEPWALQILALITLALPYQARKKTFPSVHEVNAVENYANDSLQMTDQYNKESDD 60  
 DB 1 MAEPWALQILALITLALPYQARKKTFPSVHEVNAVENYANDSLQMTDQYNKESDD 60

QY 61 KYHRIIRVLYKQROVTDHLEHYHLNVEMQWTTCKPRTTNCVPORELEHKQVNCFSVFA 120  
 DB 61 KYHRIIRVLYKQROVTDHLEHYHLNVEMQWTTCKPRTTNCVPORELEHKQVNCFSVFA 120

QY 121 VPWFQYKILNKSCSSD 137  
 DB 121 VPWFQYKILNKSCSSD 137

RESULT 2  
 CS11\_MOUSE STANDARD; PRT; 139 AA.  
 AC Q9D269; 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DE 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Cystatin 11 precursor.  
 GN CST11.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Epidiymis;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Atakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 RA Aikawa K., Iwata M., Nishi K., Kiyosawa H., Kando S., Yamanka I.,  
 RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
 RA Kiehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Blake J., Boffelli D., Fujuno M., Aono H., Baldarelli R., Barish G.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Guetlinich S., Hill D., Hotmann W., Hume D.A., Kamuya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyshew-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohseki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 CC -1- SUBCELLULAR LOCATION: Secreted (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE CYSTATIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; AK020300; BAB32061.1; -  
 DR HSSP; P01034; 1G96.  
 DR MGD; MGI:1925490; Cst11.  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF00031; Cystatin; 1.  
 DR SMART; SM00043; CY; 1.  
 DR PROSITE; PS00287; CYSTATIN; FALSE NEG.  
 KM Thiol protease inhibitor; Signal.  
 FT SIGNAL 1 28  
 FT CHAIN 29 139  
 FT SITE 76 80  
 FT DISULFID 94 102  
 FT DISULFID 115 135  
 FT CARBOHYD 134 134  
 FT SEQUENCE 139 AA; 16217 MW; F228D9815FA32640 CRC64;  
 SQ

Query March 57.3%; Score 422; DB 1; Length 139;  
 Best Local Similarity 54.5%; Pred. No. 1.2e-34;  
 Matches 73; Conservative 33; Mismatches 28; Indels 0; Gaps 0;

QY 2 AEPWALQILALITLALPYQARKKTFPSVHEVNAVENYANDSLQMTDQYNKESDDK 61  
 DB 3 AGSWKATRLILALITLALVAVSYQVKRTFIRIEVSALSSVKETLTVYDEVKESDD 62

QY 62 YHRIIRVLYKQROVTDHLEHYHLNVEMQWTTCKPRTTNCVPORELEHKQVNCFSVFAV 121  
 DB 63 YHRIIRVLYKQROVTDHLEHYHLNVEMQWTTCKPRTTNCVPORELEHKQVNCFSVFAV 122

QY 122 VPWFQYKILNKSCS 135  
 DB 123 VPWFQYKILNKSCS 136

RESULT 3  
 CST8\_MOUSE STANDARD; PRT; 142 AA.  
 AC P32766; 089102; 01-OCT-1993 (Rel. 27, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cystatin-related epididymal spermatozoal protein precursor (Cystatin-  
 DE related epididymal specific protein) (Cystatin 8).  
 GN CST8 OR CRES.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C3H, and CD-1;  
 RX MEDLINE=99247899; PubMed=10229662;  
 RA Cornwall G.A., Heia N., Sutton H.G.;  
 RT "Structure, alternative splicing and chromosomal localization of the  
 RT Cystatin-related epididymal spermatozoal gene";  
 RL Biochem. J. 340:85-93(1999).  
 RN [2]  
 RP SEQUENCE OF 4-142 FROM N.A.

CC TISSUE=Epithelium;  
 RX MEDLINE=93078799; PubMed=1280328;  
 RA Cornwell G.A., Orgebin-Crist M.-C., Hann S.R.;  
 RT "The CRIS gene: a unique testis-regulated gene related to the cystatin  
 family is highly restricted in its expression to the proximal region  
 of the mouse epididymis";  
 RL Mol. Endocrinol. 6:1653-1664(1992).  
 CC -1- FUNCTION: PERFORMS A SPECIALIZED ROLE DURING SPERM DEVELOPMENT AND  
 CC -1- MATURATION.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: PROXIMAL CAPUT REGION OF THE EPIDIDYMIS. LOWER  
 CC EXPRESSION IN THE TESTIS. WITHIN THE TESTIS IT IS LOCALIZED TO THE  
 CC ELONGATING SPERMATIDS, WHEREAS WITHIN THE EPIDIDYMIS IT IS  
 CC EXCLUSIVELY SYNTHESIZED BY THE PROXIMAL CAPUT EPITHELIUM.  
 CC -1- INDUCTION: TESTICULAR FACTORS OR HORMONES OTHER THAN ANDROGENS  
 CC PRESENT IN THE TESTICULAR FLUID MAY BE INVOLVED IN THE REGULATION  
 CC OF CRIS GENE EXPRESSION.  
 CC -1- SIMILARITY: BELONGS TO THE CYSTATIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF091503; AAC61754.1; -  
 DR EMBL: AF090691; AAC6316.1; -  
 DR EMBL: S49926; AAC3590.1; -  
 DR PIR: A45361; A45361.  
 DR HSSP: P01034; 1G96.  
 DR MGD: MGI:107161; Ccr8.  
 DR InterPro: IPR000010; Cystatin.  
 DR Pfam: PF00031; Cystatin; 1.  
 DR SMART: SM00043; CY; 1.  
 DR K101 protease inhibitor; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 142  
 FT SITE 77 81 SECONDARY AREA OF CONTACT (POTENTIAL).  
 FT DISULFID 95 105 BY SIMILARITY.  
 FT DISULFID 119 139 BY SIMILARITY.  
 FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 4 15 PLMSTLFIIP -> GRDRQVGSQK (IN REF. 2).  
 SO SEQUENCE 142 AA; 16288 MW; 50B446B9F6673B CRC64;  
 Query Match 30.8%; Score 226.5; DB 1; Length 142;  
 Best Local Similarity 35.1%; Pred. No. 1.9e-15;  
 Matches 52; Conservative 34; Mismatches 39; Indels 23; Gaps 7;  
 QY 1 MAFK-NQALQLLILTLTLPYQARKKTFLSVHEWAVN-----AKDSLQW 49  
 DB 1 MAFKLM--LSLTLTIPVLAALVGVDSK-----NEVVA--ONYGASINISANVQCVWF 51  
 QY 50 ITDYNKESDQKRYFRFLRVLVKQVQVTLHLYHLMVEMQWTCQKP--ETTNVCVQGR- 106  
 DB 52 AMEYNESSEBKRYFLVDKILHAKLQITDRMEYQIDVQISNCKKPLANTENNCIPQKKP 111  
 QY 107 ELHKVNCPSVPAVPEFYKILNKC 134  
 DB 112 ELERKMSCSFLVGLALPNNGEFNLISKEC 139  
 RESULT 4  
 CST8 HUMAN STANDARD; PRT; 142 AA.  
 AC 060676;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cystatin-related epididymal spermatogenic protein precursor (Cystatin

DE 8).  
 GN CST8 OR CRIS.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=95344753; PubMed=7619504;  
 RA Cornwell G.A., Hann S.R.;  
 RT "Transient appearance of CRIS protein during spermatogenesis and  
 RT caput epididymal sperm maturation";  
 RL Mol. Reprod. Dev. 41:37-46(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21638749; PubMed=11780052;  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
 RA Beasley O.P., Bird C.P., Blakey S.B., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,  
 RA Ellington A.G., Frankland J.A., Frazer A., French L., Garner P.,  
 RA Grahame D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Lenvassatho M.H., Levertha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McConachie L.J., McElay K., McNulty A.A.,  
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Rose M.T., Scott C.B., Sehra H.K., Shownkeen R., Sims S.,  
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulton J.E.,  
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Williams S.A.,  
 RA Whitehead S.L., Whiteaker P., Willey D.L., Williams L., Williams S.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).  
 CC -1- FUNCTION: PERFORMS A SPECIALIZED ROLE DURING SPERM DEVELOPMENT AND  
 CC MATURATION.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: PROXIMAL CAPUT REGION OF THE EPIDIDYMIS. LOWER  
 CC EXPRESSION IN THE TESTIS. WITHIN THE TESTIS IT IS LOCALIZED TO THE  
 CC ELONGATING SPERMATIDS, WHEREAS WITHIN THE EPIDIDYMIS IT IS  
 CC EXCLUSIVELY SYNTHESIZED BY THE PROXIMAL CAPUT EPITHELIUM.  
 CC -1- SIMILARITY: BELONGS TO THE CYSTATIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF059244; AAC14707.1; -  
 DR EMBL: AL109954; CAB64234.1; -  
 DR HSSP: P01034; 1G96.  
 DR Genew: HGNC:2480; CST8.  
 DR InterPro: IPR000010; Cystatin.  
 DR GO: GO:0004669; F:cysteine protease inhibitor activity; TAS.  
 DR Pfam: PF00031; Cystatin; 1.  
 DR SMART: SM00043; CY; 1.  
 DR K101 protease inhibitor; Signal; Polymorphism.  
 FT SIGNAL 1 21  
 FT CHAIN 22 142  
 FT SITE 22 142 CYSSTATIN-RELATED EPIDIDYMAL SPERMATOGENIC  
 FT PROTEIN.

FT SITE 77 81 SECONDARY AREA OF CONTACT (POTENTIAL).  
 FT DISULFID 95 105 BY SIMILARITY.  
 FT DISULFID 119 139 BY SIMILARITY.  
 FT CARBOHYD 27 27 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 39 39 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 142 142 A-> P (IN dBSNP:1054633).  
 FT SEQUENCE 142 AA; 16275 MW; 9A3512757E0FAEDC CRC64;  
 SQ SEQUENCE 142 AA; 16275 MW; 9A3512757E0FAEDC CRC64;  
 Query Match 27.4%; Score 202; DB 1; Length 142;  
 Best Local Similarity 36.3%; Pred. No. 4.9e-13;  
 Matches 49; Conservative 29; Mismatches 51; Indels 6; Gaps 4;  
 QY 5 WQALQLLAILLTLMALPYQARKKT--FLSYHEWVAWVAVAKDSIQWITDQYNKESDQY 62  
 DB 6 WLSL-ILITFLVALVARDPKQKGTGVARKLPVNASANKQCLMFAMQSYNESEDKY 64  
 QY 63 HFRIFRYLVKQRYVTHLEHNLVEMQNTTQCPETTN--CVQPER-ELHKVQCFPSVF 119  
 DB 65 VFLVVKTLQAOQVLTNLEVLIDVEIARSDCKRPLSTNEICAIQDENSEKRLKSLCSFLVG 124  
 QY 120 AVFWPEQYKILNKSC 134  
 DB 125 ALPMNGEFTVWEKXC 139  
 RESULT 5  
 CSTR\_RAT STANDARD; PRT; 142 AA.  
 ID CSTR\_RAT STANDARD; PRT; 142 AA.  
 AC 088969;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Cystatin-related epididymal spermatogenic protein precursor (Cystatin  
 8).  
 GN CSTR OR CRES.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Epididymis;  
 RX MEDLINE=99247899; PubMed=10229662;  
 RA Cornwall G.A., Hala N., Sutton H.G.;  
 RT "Structure, alternative splicing and chromosomal localization of the  
 RL Cystatin-related epididymal spermatogenic gene.";  
 RL Biochem. J. 340:85-93(1999).  
 CC -1- FUNCTION: PERFORMS A SPECIALIZED ROLE DURING SPERM DEVELOPMENT AND  
 CC MATURATION.  
 CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE CYSTATIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 CC DR EMBL: AF090692; AAC36317.1; -  
 CC DR HSSP: P01034; 1G96.  
 CC DR InterPro: IPR000010; Cystatin.  
 CC DR Pfam: PF00031; Cystatin; 1.  
 CC DR SMART: SM00043; CY; 1.  
 CC KM Thiol protease inhibitor; Signal.  
 FT SIGNAL 1 19 POTENTIAL.  
 FT CHAIN 20 142 CYSTATIN-RELATED EPIDIDYMAL SPERMATOGENIC  
 FT PROTEIN.  
 FT SITE 77 81 SECONDARY AREA OF CONTACT (POTENTIAL).  
 FT DISULFID 95 105 BY SIMILARITY.  
 FT DISULFID 119 139 BY SIMILARITY.

FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 142 AA; 16246 MW; FB873FAA6B6CAB34 CRC64;  
 Query Match 27.4%; Score 202; DB 1; Length 142;  
 Best Local Similarity 32.4%; Pred. No. 4.9e-13;  
 Matches 45; Conservative 30; Mismatches 46; Indels 18; Gaps 5;  
 QY 8 LQLLAILLTLMALPYQARKKTFLSYHEWVAWVAVAKDSIQWITDQYNKES 58  
 DB 7 LSLIFLIIPALAVDVQSK-----NEVKAQRFSGISISNNVVKQCFAMKEYNKGS 60  
 QY 59 DDKTHFRIFRYLVKQRYVTHLEHNLVEMQNTTQCP--ETTNVQPER-ELHKVQCF 115  
 DB 61 EDKTLFLDIDTHTLHATLQITPRMEYHIDVQISRSNCRPLNNTEICIPQKPKLEKULCS 120  
 QY 116 FSVFVFWPEQYKILNKSC 134  
 DB 121 FLVGALPMNGEFTVWEKXC 139  
 RESULT 6  
 CYTC\_MOUSE STANDARD; PRT; 140 AA.  
 ID CYTC\_MOUSE STANDARD; PRT; 140 AA.  
 AC P21460;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Cystatin C precursor (Cystatin 3).  
 GN CSTR3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BA1B/C; TISSUE=Brain;  
 RX MEDLINE=91054522; PubMed=2241983;  
 RA Solem M., Rawson C., Lindburg K., Barnes D.;  
 RT "Transforming growth factor beta regulates cystatin C in serum-free  
 RT mouse embryo (SFM) cells.";  
 RL Biochem. Biophys. Res. Commun. 172:945-951(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SV; TISSUE=Liver;  
 RX MEDLINE=95137392; PubMed=7835704;  
 RA Huh C., Nagle J.W., Kozak C.A., Abrahamson M., Karlsson S.;  
 RT "Structural organization, expression and chromosomal mapping of the  
 RL mouse cystatin-C-encoding gene (Cstr3).";  
 RL Gene 152:221-226(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1LS, and ISS.  
 RX MEDLINE=21363810; PubMed=11471062;  
 RA Edlinger M.A., Thompson J., Conroy O., Xu Y., Yang F., Cammiff J.,  
 RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikeja J.M.;  
 RT "High-throughput sequence identification of gene coding variants  
 RT within alcohol-related QTLs";  
 RL Mamm. Genome 12:657-663(2001).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,  
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udén T.B., Tosihyoki S., Carninci P., Prange C.,  
 RA Rana S.S., Loggiano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
 RA Bosak S.A., McEwan P.U., McKernan K.J., Malek J.A., Gunnarsson P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.U., Huijck S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmitz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Small D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).  
 CC -1- FUNCTION: AS AN INHIBITOR OF CYSTEINE PROTEINASES, THIS PROTEIN IS  
 CC THOUGHT TO SERVE AN IMPORTANT PHYSIOLOGICAL ROLE AS A LOCAL  
 CC REGULATOR OF THIS ENZYME ACTIVITY.  
 CC -1- SIMILARITY: BELONGS TO THE CYSTATIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR EMBL; M59470; AAA63298.1; -;  
 DR EMBL; U10098; AAB41056.1; -;  
 DR EMBL; AF483486; AAL90760.1; -;  
 DR EMBL; AF483487; AAL90761.1; -;  
 DR EMBL; BC002072; AALH02072.1; -;  
 DR PIR; A36163; A36163.  
 DR HSSP; P01034; I596.  
 DR MGD; MGI:102519; Ccst3.  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF00031; Cystatin; 1.  
 DR SMART; SM00043; CY; 1.  
 DR PROSITE; PS00287; CYSTATIN; 1.  
 KM Thiol protease inhibitor; Signal.  
 FT SIGNAL 1 20  
 FT CHAIN 21 140  
 FT ACT\_SITE 31 31  
 FT SITE 75 79  
 FT DISULFID 93 103  
 FT DISULFID 117 137  
 FT CONFLICT 16 16  
 FT CONFLICT 84 84  
 FT SEQUENCE 140 AA; 15531 MW; 3A563406D58D0F5 CRC64;  
 SQ

Query Match 25.3%; Score 186.5; DB 1; Length 140;  
 Best Local Similarity 30.9%; Pred. No. 1.6e-11;  
 Matches 43; Conservative 30; Mismatches 63; Indels 3; Gaps 2;

QY 1 MAEFQALQLLILTLALPYQARKTFLSVHEVMAVENYAKDSLQWITDQYNKESDD 60  
 DB 1 MASPLRSILFLAVLAVALMAATPKQGPRLGAPEDADNEGVRALDPFAVSEYKSGSND 60

QY 61 KYHRIPIRVLVKQVQVTDHLEHVLNEMQWTTQCKPEP--TNC--VPOBERLHKVNCPEFS 117  
 DB 61 AYHSRAIQVVARQVLAAGVAVFLDVEGRITCTKQTNLTDCPHDQPHLMRLALCSFQ 120

QY 118 VFAVPWFQYKILNKSCSS 136  
 DB 121 IYSVPWKQTHSLTKFSCKN 139

RESULT 7  
 CYTC\_RABIT STANDARD; PRT; 148 AA.  
 AC 097862;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cystatin C precursor.  
 GN CSTR3.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Japanese White; TISSUE=Bone;  
 RX MEDLINE=98424349; PubMed=9753427;  
 RA Kobori M., Ikeda Y., Nara H., Kato M., Kamegawa M., Nojima H.,  
 RA Kawahima H.,  
 RT "Large scale isolation of osteoclast-specific genes by an improved  
 method involving the preparation of a subtracted cDNA library."  
 RL Genes Cells 3:459-475(1998).  
 CC -1- FUNCTION: THIS IS A THIOL PROTEINASE INHIBITOR.  
 CC -1- SIMILARITY: BELONGS TO THE CYSTATIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR EMBL; AB009342; BAA75921.1; -;  
 DR HSSP; P01034; I596.  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF00031; Cystatin; 1.  
 DR SMART; SM00043; CY; 1.  
 DR PROSITE; PS00287; CYSTATIN; FALSE\_NEG.  
 KM Thiol protease inhibitor; Signal.  
 FT SIGNAL 1 28  
 FT CHAIN 29 148  
 FT ACT\_SITE 39 39  
 FT SITE 83 87  
 FT DISULFID 101 111  
 FT DISULFID 125 145  
 FT SEQUENCE 148 AA; 16346 MW; 1523C831169E5B9A CRC64;  
 SQ

Query Match 25.1%; Score 184.5; DB 1; Length 148;  
 Best Local Similarity 30.4%; Pred. No. 2.7e-11;  
 Matches 42; Conservative 34; Mismatches 51; Indels 11; Gaps 4;

QY 10 LLLALILTLNALPY-----QARKTFL--SVHEVMAVENYAKDSLQWITDQYNKESDDK 61  
 DB 10 LLLAALVVALALNAPPAAGRTSPRLGLGELVDAQDEKQVQALGPAEYKSGSND 69

QY 62 YHRIPIRVLVKQVQVTDHLEHVLNEMQWTTQCKPEP--TNC--VPOBERLHKVNCPEFSV 118  
 DB 70 YHSRALQVVARQVLAAGVAVFLDVEGRITCTKQTNLTDCPHDQPHLMRLALCSFQ 129

QY 119 VFAVPWFQYKILNKSCSS 136  
 DB 130 YSVPMKISILKSDCN 147

RESULT 8  
 CYTC\_SAISC STANDARD; PRT; 146 AA.  
 AC 019093;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cystatin C precursor.  
 GN CSTR3.  
 OS Saimiri sciureus (Common squirrel monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.  
 NCBI\_TaxID=9521;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97054523; PubMed=8898820;  
 RA Wei L.H., Walker L.C., Levy B.,  
 RT "Cystatin C. Icelandic-like mutation in an animal model of  
 RT cerebrovascular beta-amyloidosis."  
 RL Stroke 27:2080-2085(1996).

-1- FUNCTION: AS AN INHIBITOR OF CYSTEINE PROTEINASES, THIS PROTEIN IS THOUGHT TO SERVE AN IMPORTANT PHYSIOLOGICAL ROLE AS A LOCAL REGULATOR OF THIS ENZYME ACTIVITY.

-1- SIMILARITY: BELONGS TO THE CYSTATIN FAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL: U52028; AAB64051.1; -.

DR HSSP: P01034; 1G96.

DR InterPro: IPR000010; Cystatin.

DR Pfam: PF00031; cystatin.1.

DR SMART: SM00043; CY, 1.

DR PROSITE, PS00287; CYSTATIN; 1.

KM Thiol protease inhibitor; Amyloid; Signal.

FT SIGNAL 1 26 BY SIMILARITY.

FT CHAIN 27 146 CYSTATIN C.

FT ACT SITE 37 37 REACTIVE SITE.

FT SITE 81 85 SECONDARY AREA OF CONTACT.

FT DISULFID 99 109 BY SIMILARITY.

FT DISULFID 123 143 BY SIMILARITY.

SQ SEQUENCE 146 AA; 15946 MW; 08196353C0306AA3 CRC64;

Query March 24.6%; Score 181; DB 1; Length 146;  
 Best Local Similarity 34.2%; Pred. No. 5.8e-11;  
 Matches 50; Conservative 24; Mismatches 62; Indels 10; Gaps 5;

QY 1 MAEPQALQLLALTLALPYQA-----RKTEPL-SVHEWMAVENAKDSLQWITQY 54  
 Db 1 MAGELRAPLLALIALAVALASPAGAPRGTRPRLGSPMNASVEEGVRLADPAVSEY 60

QY 55 NKESDDKHYFRIFRLKYQGVDTLLEHYNVEMQWTCOK--PETTNCVPOER-ELRKQ 111  
 Db 61 NKAANDMYHSRALQVVRARKQIVAGVNVFLDVEMGRTTCTGNQRPDLDCPFHEQPHLRK 120

QY 112 VNCFSVPVAVPFEQYKINKSCSD 137  
 Db 121 AFGSFQIYSVPV-QGIMTLKSTCOD 145

RESULT 9  
 CYTT\_HUMAN STANDARD; PRT; 141 AA.  
 AC P09228; Q9UCQ7;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cystatin SA precursor (Cystatin S5).  
 GN CST2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88185836; PubMed=3446578;  
 RA Satoh E., Kim H.-S., Smithies O., Maeda N.;  
 RT "Human cysteine-proteinase inhibitors: nucleotide sequence analysis  
 of three members of the cystatin gene family.";  
 RL Gene 61:329-338(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21638749; PubMed=11780052;  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggley C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,

RA Chapman J.C., Ciamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clough S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,  
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.T., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Levaesalho M.H., Leverisa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA March V.L., Martin S.L., McConachie I.J., McElay K., McMurtry A.A.,  
 RA Milne S.A., Mistry D., Moore M.J.P., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,  
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Williams S.A.,  
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).

[3]  
 RP SEQUENCE OF 21-40.  
 RC TISSUE=Saliva;  
 RX MEDLINE=92138674; PubMed=1778989;  
 RA Isemura S., Satoh E., Sanada K., Minakata K.;  
 RT "Characterization and amino acid sequence of a new acidic cysteine  
 proteinase inhibitor (cystatin SA) structurally closely related to  
 cystatin S, from human whole saliva.";  
 RL J. Biochem. 110:648-654(1991).

[4]  
 RP SEQUENCE OF 25-141.  
 RX MEDLINE=88139220; PubMed=3436950;  
 RA Isemura S., Satoh E., Sanada K.;  
 RT "Characterization and amino acid sequence of a new acidic cysteine  
 proteinase inhibitor (cystatin SA) structurally closely related to  
 cystatin S, from human whole saliva.";  
 RL J. Biochem. 102:693-704(1987).

[5]  
 RP PRELIMINARY SEQUENCE OF 25-141.  
 RA Isemura S., Satoh E., Sanada K., Isemura M., Ito S.;  
 RT "Characterization and amino acid sequence of a new acidic cysteine  
 proteinase inhibitor (cystatin SA) structurally closely related to  
 cystatin S, from human whole saliva.";  
 RL (in) Turk V. (eds.);  
 RL Cysteine proteinases and their inhibitors, pp. 497-505,  
 RL Walter de Gruyter, Berlin and New York (1986).

[6]  
 RP SEQUENCE OF 25-141 FROM N.A.  
 RX MEDLINE=89076505; PubMed=3202964;  
 RA Satoh E., Isemura S., Sanada K., Kim H.-S., Smithies O., Maeda N.;  
 RT "Cystatin superfamily. Evidence that family II cystatin genes are  
 evolutionarily related to family III cystatin genes.";  
 RL Biol. Chem. Hoppe-Seyler 369:191-197(1988).

CC -1- FUNCTION: Thiol protease inhibitor.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE CYSTATIN FAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

CC EMBL: M19673; AAA36116.1; -.

DR EMBL: M19671; AAA36116.1; JOINED.

DR EMBL: M19672; AAA36116.1; JOINED.

DR EMBL: AL591074; CAC94784.1; -.

DR PIR: B29632; B29632.

DR HSSP: P01034; 1G96.  
 DR Genew, HGNC:2474; CST2.  
 DR MTM; 123856; --  
 DR GO; GO:0004869; F:cysteine protease inhibitor activity; TAS.  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF00031; cystatin; 1.  
 DR SMART; SM00043; CY; 1.  
 DR PROSITE; PS00287; CYSTATIN; 1.  
 DR Thiol protease inhibitor; Signal; Multigene family.  
 KW Thiol protease inhibitor; Signal; Multigene family.  
 FT SIGNAL 1 20  
 FT CHAIN 21 141  
 FT ACT SITE 32 32 REACTIVE SITE.  
 FT SITE 76 80 SECONDARY AREA OF CONTACT.  
 FT DISULFID 94 104 BY SIMILARITY.  
 FT DISULFID 118 138 BY SIMILARITY.  
 SQ SEQUENCE 141 AA; 16445 MW; BB54915B1B977AA2 CRC64;  
 Query Match 23.6%; Score 174; DB 1; Length 141;  
 Best Local Similarity 25.7%; Pred. No. 2.7e-10;  
 Matches 35; Conservative 36; Mismatches 59; Indels 6; Gaps 3;  
 QY 5 WQALQLLALILTLMLPYQARKKTP---SVHEVAVENVAKDSLQMTDQNKESDK 61  
 DB 3 WPLCTLLLATQVALWSPQSDRIIEGIVDADLDERVGRALHPVISEVKKATEDE 62  
 QY 62 YHRIPIVAVKQROVTDHLBYHLVENVQMTTCQ--PETNVCVQER-ELHKVNCFFSV 118  
 DB 63 YHRIPIVAVKQROVTDHLBYHLVENVQMTTCQ--PETNVCVQER-ELHKVNCFFSV 118  
 QY 119 FAVPMFQYKILKSC 134  
 DB 123 YEVPMEDRMVLSVRS 138  
 RESULT 10  
 CYT CHICK STANDARD; PRT; 139 AA.  
 AC P01038;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cystatin precursor (Egg-white cystatin).  
 OS Gallus gallus (Chicken).  
 OC Eukaryote; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OC NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9000873; Pubmed=2793849;  
 RA Colella R., Sakaguchi Y., Nagase H., Bird J.W.C.;  
 RT "Chicken egg white cystatin. Molecular cloning, nucleotide sequence,  
 RT and tissue distribution.";  
 RL J. Biol. Chem. 264:17164-17169 (1989).  
 RN [2]  
 RP SEQUENCE OF 24-139.  
 RX MEDLINE=94178305; Pubmed=6712597;  
 RA Schwabe C., Anastasi A., Crow H., McDonald J.R., Barrett A.J.;  
 RT "Cystatin. Amino acid sequence and possible secondary structure.";  
 RL Biochem. J. 217:813-817 (1984).  
 RN [3]  
 RP SEQUENCE OF 24-139.  
 RX MEDLINE=94110059; Pubmed=6662498;  
 RA Turk V., Brzin J., Longer M., Ritonja A., Bropink M., Borchart U.,  
 RT "Protein inhibitors of cysteine proteinases. III. Amino-acid sequence  
 RT of cystatin from chicken egg white.";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 364:1487-1496 (1983).  
 RN [4]  
 RP CHARACTERIZATION OF PROTEIN.  
 RX MEDLINE=93256421; Pubmed=6409085;  
 RA Anastasi A., Brown M.A., Kembhavi A.A., Nicklin M.J.H., Sayers C.A.,  
 RA Sunter D.C., Barrett A.J.;

RT "Cystatin, a protein inhibitor of cysteine proteinases. Improved  
 RT purification from egg white, characterization, and detection in  
 RT chicken serum.";  
 RL Biochem. J. 211:129-138 (1983).  
 RN [5]  
 RP DISULFIDE BONDS.  
 RA Grubb A., Loeffberg H., Barrett A.J.;  
 RT "The disulphide bridges of human cystatin C (gamma-trace) and chicken  
 RT cystatin.";  
 RL FEBS Lett. 170:370-374 (1984).  
 RN [6]  
 RP PHOSPHORYLATION.  
 RX MEDLINE=89252033; Pubmed=2721673;  
 RA Laber B., Krieglstein K., Henschen A., Kos J., Turk V., Huber R.,  
 RT "The cysteine proteinase inhibitor chicken cystatin is a  
 RT phosphoprotein.";  
 RL FEBS Lett. 248:162-168 (1989).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).  
 RX MEDLINE=89052676; Pubmed=3191914;  
 RA Bode W., Engh R., Musil D., Thiele U., Huber R., Karshkov A.,  
 RA Brzin J., Kos J., Turk V.;  
 RT "The 2.0 A X-ray crystal structure of chicken egg white cystatin and  
 RT its possible mode of interaction with cysteine proteinases.";  
 RL EMBO J. 7:2593-2599 (1988).  
 RN [8]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE=94087719; Pubmed=8263912;  
 RA Dieckmann T., Mitschang L., Hofmann M., Kos J., Turk V.,  
 RA Auerwald R.A., Jesenicek R., Oschkinat H.;  
 RT "The structures of native phosphorylated chicken cystatin and of a  
 RT recombinant unphosphorylated variant in solution.";  
 RL J. Mol. Biol. 234:1048-1059 (1993).  
 CC -1- FUNCTION: THIS PROTEIN BINDS TIGHTLY TO AND INHIBITS A VARIETY OF  
 CC THIOL PROTEASES INCLUDING P1CIN, PAPAIN, AND CATHEPSINS B, C, H,  
 CC AND L. ALTHOUGH ISOLATED FROM EGG WHITE, IT IS ALSO PRESENT IN  
 CC SERUM.  
 CC -1- SIMILARITY: BELONGS TO THE CYSTATIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: J05077; AAA48744.1; --  
 DR FIR; A34456; UDCH.  
 DR PDB; 1CEW; 31-JAN-94.  
 DR PDB; 1A67; 27-MAY-98.  
 DR PDB; 1A90; 17-JUN-98.  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF00031; cystatin; 1.  
 DR SMART; SM00043; CY; 1.  
 DR PROSITE; PS00287; CYSTATIN; 1.  
 KW Thiol protease inhibitor; Phosphorylation; Signal; 3D-structure.  
 FT SIGNAL 1 23  
 FT CHAIN 24 139  
 FT ACT SITE 32 32 CYSTATIN.  
 FT SITE 76 80 REACTIVE SITE.  
 FT DISULFID 94 104 SECONDARY AREA OF CONTACT.  
 FT DISULFID 118 138  
 FT MOD RES 103 103  
 FT STRAND 35 36 PHOSPHORYLATION (PARTIAL).  
 FT TURN 39 40  
 FT TURN 42 41  
 FT HELIX 42 51  
 FT TURN 52 52  
 FT HELIX 53 56  
 FT TURN 57 58  
 FT STRAND 63 77  
 FT STRAND 81 95

```

FT TURN 96 97
FT TURN 99 100
FT HELIX 101 108
FT STRAND 115 125
FT TURN 126 129
FT STRAND 130 139
SQ SEQUENCE 139 AA; 15287 MW; D92D1131C4D37891 CRC64;

Query Match 23.6%; Score 173.5; DB 1; Length 139;
Best Local Similarity 33.1%; Pred. No. 3e-10;
Matches 43; Conservative 25; Mismatches 57; Indels 5; Gaps 3;

QY 10 LLLAILETLMALPQARKKFLSYHEWVAVEN--YAKDSLQWITDQYNKESDDKXHFRI 67
DB 9 VLLAALMTVGAVLGSEDRSLGLAPVVDENDEGLQALQPAAEYVRASNDKXSRV 68
QY 68 RVLKVGQVTDHLEHNLVEMQWTTQKP--ETTNC-VPOERELHKQVCFPSVPAVWF 124
DB 69 RVISAKQLVSGIKYILQVEIGRTTCFKSGDLQSCFHFDEPEMAKYTTCTFVYVSI 128
QY 125 EGYKILKSC 134
DB 129 NQIKLESKC 138

RESULT 11
CYTC_RAT STANDARD; PRT; 127 AA.
ID CYTC_RAT STANDARD; PRT; 127 AA.
AC P14841;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 26-FEB-2003 (Rel. 41, Last annotation update)
DE Cystatin C precursor (Fragment).
GN CST3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=Buffalo;
RX MEDLINE=90092122; PubMed=2689174;
RA Cole T., Dickson P.W., Esmard F., Averill F., Riedinger G.,
RA Gauthier F., Schreiber G.;
RT "The cDNA structure and expression analysis of the genes for the
RT cysteine proteinase inhibitor cystatin C and for beta 2-microglobulin
RT in rat brain.";
RL Eur. J. Biochem. 186:35-42(1989).
RN 12
RP SEQUENCE OF 8-127.
RX MEDLINE=90380276; PubMed=2400577;
RA Esmard F., Esmard A., Faucher D., Capony J.-P., Derancourt J.,
RA Brillard M., Gauthier F.;
RT "Rat cystatin C: the complete amino acid sequence reveals a site for
RT N-glycosylation.";
RL Biol. Chem. Hoppe-Seyler 371:161-166(1990).
RN 13
RP SEQUENCE OF 8-49.
RX MEDLINE=88313020; PubMed=3044831;
RA Esmard A., Esmard F., Guillou F., Gauthier F.;
RT "Two rat homologues of human cystatin C.";
RL FEBS Lett. 236:475-478(1988).
RN 14
RP SEQUENCE OF 8-20.
RC TISSUE=Sertoli cells;
RX MEDLINE=9225121; PubMed=1563513;
RA Esmard A., Esmard F., Guillou F., Gauthier F.;
RT "Production of the cysteine proteinase inhibitor cystatin C by rat
RT Sertoli cells.";
RL FEBS Lett. 300:131-135(1992).
CC -1- FUNCTION: AS AN INHIBITOR OF CYSTEINE PROTEINASES, THIS PROTEIN IS
CC THOUGHT TO SERVE AN IMPORTANT PHYSIOLOGICAL ROLE AS A LOCAL
CC REGULATOR OF THIS ENZYME ACTIVITY. KNOWN TO INHIBITS CATHEPSIN B,

```

```

CC H, AND L.
CC -1- SIMILARITY: BELONGS TO THE CYSTATIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; X16957; CAA34831.1; -.
CC PIR; S07085; S07085.
CC HSSP; P01034; 1G96.
CC InterPro; IPR000010; Cystatin.
CC Pfam; PF00031; cystatin; 1.
CC SMART; SM00043; Cy; 1.
CC PROSITE; PS00287; CYSTATIN; 1.
CC K101 protease inhibitor; Signal.
CC NON TER 1
CC SIGNAL 1
CC CHAIN 8 127 CYSTATIN C.
CC ACT_SITE 18 18 REACTIVE SITE.
CC SITE 62 66 SECONDARY AREA OF CONTACT.
CC DISULFID 80 90 BY SIMILARITY.
CC DISULFID 104 124 BY SIMILARITY.
CC CONFLICT 25 25 A -> E (IN REF. 2).
SQ SEQUENCE 127 AA; 14039 MW; 78F70158B7925853 CRC64;

Query Match 23.3%; Score 171.5; DB 1; Length 127;
Best Local Similarity 33.3%; Pred. No. 4.3e-10;
Matches 35; Conservative 24; Mismatches 43; Indels 3; Gaps 2;

QY 35 EWMAVENYAKDSLQWITDQYNKESDDKXHFRI RYLVKVGQVTDHLEHNLVEMQWTTQ 94
DB 22 EADASEGVGRALDFAVSEFNKGSNDYHSRAIQVAPARQLVAGIYVYLDVEMGRITCT 81
QY 95 KPET--TNC-VPOERELHKQVCFPSVPAVWFQYKILKSCS 136
DB 82 KSGTNLTNCFPHDQPHLMKALCSFOIYSVPWKGTHRLTSSGCA 126

RESULT 12
CYTC_HUMAN STANDARD; PRT; 146 AA.
ID CYTC_HUMAN STANDARD; PRT; 146 AA.
AC P01034;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cystatin C precursor (Neuroendocrine basic polypeptide) (Gamma-trace)
DE (Post-gamma-globulin).
GN CST3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE=Leukocyte;
RX MEDLINE=87219149; PubMed=2495457;
RA Abrahamson M., Grubb A., Olafsson I., Lundwall A.;
RT "Molecular cloning and sequence analysis of cDNA coding for the
RT precursor of the human cysteine proteinase inhibitor cystatin C.";
RL FEBS Lett. 216:229-233(1987).
RN 12
RP SEQUENCE FROM N.A.
RC TISSUE=Leukocyte;
RX MEDLINE=90303202; PubMed=2363674;
RA Abrahamson M., Olafsson I., Palodocir A., Ulvbaeck M., Lundwall A.,
RA Jensen O., Grubb A.;
RT "Structure and expression of the human cystatin C gene.";
RL Biochem. J. 268:287-294(1990).
RN 13

```

RP SEQUENCE FROM N.A. (HCHMA VARIANT).  
 RC TISSUE=Brain.  
 RX MEDLINE=99235594; PubMed=2541223;  
 RA Levy B., Lopez-Otin C., Ghiso J.,  
 RT "Stroke in Icelandic patients with hereditary amyloid angiopathy is  
 RT related to a mutation in the cystatin C gene, an inhibitor of  
 RT cysteine proteases.";  
 RL J. Exp. Med. 169:1771-1778(1989).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89350949; PubMed=2764935;  
 RA Saitoh E., Sabatini L.M., Eddy R.L., Shows T.B., Azen E.A.,  
 RA Iemura S., Sando K.,  
 RT "The human cystatin C gene (CST3) is a member of the cystatin gene  
 RT family which is localized on chromosome 20.";  
 RL Biochem. Biophys. Res. Commun. 162:1324-1331(1989).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Dickinson D.P., Hewett-Emmett D., Thiesse M.,  
 RT "Acquisition of complex patterns of differential expression in  
 RT epithelial cell populations during the evolution of type 2 cystatin  
 RT genes.";  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21638749; PubMed=11780052;  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
 RA Beasley O.P., Bird C.P., Blakey S.B., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burdill W.D., Butler A.P., Carder C., Carter N.P., Clee C.M.,  
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Coby N.R.,  
 RA Clegg S., Cobley V.B., Collier R.E., Connor R.E., Cordy N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,  
 RA Billington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Graffham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.B., Jerosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Lehaeslath M.H., Leverisa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McComachie L.J., McElay K., McMurtry A.A.,  
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., McMurry T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H.,  
 RA Rice C.M., Ross M.T., Scott C.B., Sehra H.K., Showkseen R., Sims S.,  
 RA Stuce C.D., Smith M.L., Soderlund C., Steward C.A., Suleston J.E.,  
 RA Swann R.M., Symmore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Williams A.,  
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.,  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain.  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Krausberg R.L., Fellingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Bhac N.K.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.B., Scheetz T.E.,  
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carrinetti P., Prange C.,  
 RA Raba S.S., Loquiano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Faley J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Buttefield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [8]  
 RP SEQUENCE OF 27-146.  
 RX MEDLINE=82222268; PubMed=6283552;  
 RA Grubb A., Loeffberg H.,  
 RT "Human gamma-trace, a basic microprotein: amino acid sequence and  
 RT presence in the adenohypophysis.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 79:3024-3027(1982).  
 RN [9]  
 RP SEQUENCE OF 27-73.  
 RX MEDLINE=84110059; PubMed=6662498;  
 RA Turk V., Brzin J., Longner M., Ritonja A., Eropkin M., Borchart U.,  
 RA Machleidt W.,  
 RT "Protein inhibitors of cysteine proteinases. III. Amino-acid sequence  
 RT of cystatin from chicken egg white.";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 364:1487-1496(1983).  
 RN [10]  
 RP SEQUENCE OF 27-76.  
 RX MEDLINE=84128015; PubMed=6365094;  
 RA Brzin J., Popovic T., Turk V.,  
 RT "Human cystatin, a new protein inhibitor of cysteine proteinases.";  
 RL Biochem. Biophys. Res. Commun. 118:103-109(1984).  
 RN [11]  
 RP DISULFIDE BONDS.  
 RA Grubb A., Loeffberg H., Barrett A.J.,  
 RT "The disulphide bridges of human cystatin C (gamma-trace) and chicken  
 RT cystatin.";  
 RL FEBS Lett. 170:370-374(1984).  
 RN [12]  
 RP X-RAY CRYSTALLOGRAPHY (2.50 ANGSTROMS) OF 27-146.  
 RX MEDLINE=21173909; PubMed=11276250;  
 RA Janowski R., Kozak M., Janowska E., Grzonka Z., Grubb A.,  
 RA Abrahamson M., Jaskolski M.,  
 RT "Human cystatin C, an amyloidogenic protein, dimerizes through  
 RT three-dimensional domain swapping.";  
 RL Nat. Struct. Biol. 8:316-320(2001).  
 RN [13]  
 RP VARIANT GLN-94.  
 RX MEDLINE=92316504; PubMed=1352269;  
 RA Abrahamson M., Jonasson S., Olafsson I., Jansson O., Grubb A.,  
 RT "Hereditary cystatin C amyloid angiopathy: identification of the  
 RT disease-causing mutation and specific diagnosis by polymerase chain  
 RT reaction based analysis.";  
 RL Hum. Genet. 89:377-380(1992).  
 CC -1- FUNCTION: AS AN INHIBITOR OF CYSTEINE PROTEINASES, THIS PROTEIN IS  
 CC THOUGHT TO SERVE AN IMPORTANT PHYSIOLOGICAL ROLE AS A LOCAL  
 CC REGULATOR OF THIS ENZYME ACTIVITY.  
 CC -1- SUBUNIT: Homodimer.  
 CC -1- TISSUE SPECIFICITY: IT IS EXPRESSED IN HIGHEST LEVELS IN THE  
 CC EPIDIDYMIS, VAS DEFERENS, BRAIN, THYMUS, AND OVARY AND THE LOWEST  
 CC IN THE SUBMANDIBULAR GLAND.  
 CC -1- DISEASE: IMPLICATED IN A HEREDITARY FORM OF CEREBRAL HEMORRHAGE  
 CC CHARACTERIZED BY A THICKENING OF THE CEREBRAL ARTERIES WALLS WITH  
 CC DEPOSITION OF MATERIAL WITH THE CHARACTERISTICS OF AMYLOID. THIS  
 CC DISEASE IS KNOWN AS HEREDITARY CEREBRAL HEMORRHAGE WITH  
 CC AMYLOIDOSIS (HCHMA), OR HEREDITARY CYSTATIN C AMYLOID ANGIOPATHY  
 CC (HCCAA) OR ALSO AS ICELANDIC CEREBROVASCULAR TYPE AMYLOIDOSIS.  
 CC -1- SIMILARITY: BELONGS TO THE CYSTATIN FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X05607; CAA29096.1; --  
 DR EMBL: X52255; CAA36497.1; --  
 DR EMBL: M27891; AAA52164.1; --

DR EMBL; M27889; AAA52164.1; JOINED.  
 DR EMBL; M27890; AAA52164.1; JOINED.  
 DR EMBL; X61681; CAA43856.2; JOINED.  
 DR EMBL; X61682; CAA43856.2; JOINED.  
 DR EMBL; X61683; CAA43856.2; JOINED.  
 DR EMBL; AF319564; AAK11570.1; -.  
 DR EMBL; AL121894; CAC05424.1; -.  
 DR EMBL; BC013083; AAK13083.1; -.  
 DR PIR; S10216; UDHU.  
 DR PDB; 1G96; 06-APR-01.  
 DR Genew; HGNC:2475; CST3.  
 DR MIM; 604312; -.  
 DR MIM; 105150; -.  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF00031; Cystatin; 1.  
 DR SMART; SM00043; CY; 1.  
 DR PROSITE; PS00287; CYSTATIN; 1.  
 DR Thiol protease inhibitor; Amyloid; signal; Disease mutation;  
 KM Polymorphism; 3D-structure.  
 FT SIGNAL 1 26  
 FT CHAIN 27 146 CYSTATIN C.  
 FT ACT\_SITE 37 37 REACTIVE SITE.  
 FT SITE 81 85 SECONDARY AREA OF CONTACT.  
 FT DISULFID 99 109  
 FT DISULFID 123 143

Query Match 23.2%; Score 171; DB 1; Length 146;  
 Best Local Similarity 31.5%; Pred. No. 5.6e-10;  
 Matches 46; Conservative 26; Mismatches 64; Indels 10; Gaps 4;

QY 1 MAEPQALQLLALITLTLALPYQARK-----KTFLSVHEWAVENYAKDSLQWITDOY 54  
 DB 1 MAGFLRPLLLILALVALAVSPAAGSPGKPRVLGSPMDASVEEGVRALDPAVSEY 60  
 QY 55 NKESDDKYHFRIPRYLVQROVTDHLEHVLNVEQMWTQCK--PETTNC-VPOREELHKQ 111  
 DB 61 NKAENDMHSRALQVBARQIVAGVNYFLDVELGRITCTQTQPNLDNCPHNDQPHLRK 120  
 QY 112 VNCFSVAVPWFPEQYKLNKSCSD 137  
 DB 121 AFGSFQIYVPM-QGTMTLSKSTCD 145

RESULT 13  
 CYTC\_MACMU STANDARD; PRT; 146 AA.  
 AC O19092;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cystatin C precursor.  
 GN CST3.  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecoidea; Macaca.  
 OX NCBI\_TaxID=9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97054523; PubMed=8898820;  
 RA Wei L.H., Walker L.C., Levy B.;  
 RT "Cystatin C. Icelandic-like mutation in an animal model of  
 RT cerebrovascular beta-amyloidosis.";  
 RL Stroke 27:2080-2085(1996).  
 CC -1- FUNCTION: AS AN INHIBITOR OF CYSTEINE PROTEINASES, THIS PROTEIN IS  
 CC THOUGHT TO SERVE AN IMPORTANT PHYSIOLOGICAL ROLE AS A LOCAL  
 CC REGULATOR OF THIS ENZYME ACTIVITY.  
 CC -1- SIMILARITY: BELONGS TO THE CYSTATIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL; U51912; AAB64050.1; -.  
 DR HSSP; P01034; 1G96.  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF00031; Cystatin; 1.  
 DR SMART; SM00043; CY; 1.  
 DR PROSITE; PS00287; CYSTATIN; 1.  
 DR Thiol protease inhibitor; Amyloid; signal.  
 KM By similarity.  
 FT SIGNAL 1 26  
 FT CHAIN 27 146 CYSTATIN C.  
 FT ACT\_SITE 37 37 REACTIVE SITE.  
 FT SITE 81 85 SECONDARY AREA OF CONTACT.  
 FT DISULFID 99 109  
 FT DISULFID 123 143 BY SIMILARITY.  
 SQ SEQUENCE 146 AA; 15857 MW; F0B3BB74A29DF26 CRC64;

Query Match 23.1%; Score 170; DB 1; Length 146;  
 Best Local Similarity 31.5%; Pred. No. 7e-10;  
 Matches 46; Conservative 26; Mismatches 64; Indels 10; Gaps 4;

QY 1 MAEPQALQLLALITLTLALPYQARK-----RKKTFLSVHEWAVENYAKDSLQWITDOY 54  
 DB 1 MAGFLRPLLLILALVALAVSPAAGSPGKPRVLGSPMDASVEEGVRALDPAVSEY 60  
 QY 55 NKESDDKYHFRIPRYLVQROVTDHLEHVLNVEQMWTQCK--PETTNCVPOER-ELHKQ 111  
 DB 61 NKAENDMHSRALQVBARQIVAGVNYFLDVELGRITCTQTQPNLDNCPHNDQPHLRK 120  
 QY 112 VNCFSVAVPWFPEQYKLNKSCSD 137  
 DB 121 AFGSFQIYVPM-QGTMTLSKSTCD 145

RESULT 14  
 CYT\_COTJA STANDARD; PRT; 116 AA.  
 AC P81061;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cystatin (Egg-white cystatin).  
 OS Coturnix coturnix japonica (Japanese quail).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Coturnix.  
 OX NCBI\_TaxID=93934;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Egg white;  
 RX MEDLINE=97420480; PubMed=9276465;  
 RA Gerhartz B., Engh R.A., Mentele R., Eckerskorn C., Torquato R.,  
 RA Wiltman J., Kolb H.J., Machleidt W., Fritz H., Auerwald B.A.;  
 RT "Quail cystatin: isolation and characterisation of a new member of  
 RT the cystatin family and its hypothetical interaction with cathepsin  
 RT B.";  
 RL FEBS Lett. 412:551-558(1997).  
 CC -1- FUNCTION: THIS PROTEIN BINDS TIGHTLY TO AND INHIBITS PAPAIN AND  
 CC CATHEPSIN B.  
 CC -----  
 CC -1- SIMILARITY: BELONGS TO THE CYSTATIN FAMILY.  
 CC -----  
 DR HSSP; P01038; ICEW.  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF00031; Cystatin; 1.  
 DR SMART; SM00043; CY; 1.  
 DR PROSITE; PS00287; CYSTATIN; 1.  
 KM Thiol protease inhibitor; Phosphorylation.  
 FT ACT\_SITE 9 9  
 FT SITE 53 57 REACTIVE SITE.  
 FT DISULFID 71 81 SECONDARY AREA OF CONTACT.  
 FT DISULFID 95 115  
 FT MOD\_RES 80 80 PHOSPHORYLATION.

SQ SEQUENCE 116 AA; 13093 MW; 46248621053A2F70 CRC64;  
 Query Match 23.0%; Score 169.5; DB 1; Length 116;  
 Best Local Similarity 35.9%; Pred. No. 6.1e-10;  
 Matches 33; Conservative 21; Mismatches 35; Indels 3; Gaps 2;  
 QY 46 SLQWITDQYNKESDDKTHFPIFVLYKQRYOTDHLFHLVNEQWTTQCK--PETTNC-V 102  
 DB 24 ALQFAMAEVYKASNDKYSRVVRILISAKQOLVSGIKYMEVEIGRTCPKSSADLOSCEF 83  
 QY 103 POEBELHKOVNCPFSVPAVPMPEQYKLINKSC 134  
 DB 84 HDEPEMAKYTTCCNPFVYSIPFWLNDIKLKSSC 115  
 RESULT 15  
 CRTL\_HUMAN STANDARD; PRT; 165 AA.  
 AC Q9H114;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cystatin-like 1 precursor.  
 GN CRTL.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 NC NCBI\_TaxID=9606;  
 RX MEDLINE=21638749; PubMed=11780052;  
 RA DeLounkas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavridis G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
 RA Bailey J., Barlow K.P., Bates K.N., Beard L.M., Beare D.M.,  
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Buttrill W.D., Butler A.P., Gardner C., Carter N.P.,  
 RA Chapman J.C., Clamp M., Collier G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.B., Collier R.E., Connor R.E., Corby N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,  
 RA Billington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Lehesvaara M.H., Leverhwa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McConachie L.J., McElay K., McMurray A.A.,  
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I., Ramsay H.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Ross M.T., Scott C.B., Sehra H.K., Showkeen R., Sims S.,  
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann R.M., Symmore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Walls J.M., Williams S.A.,  
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:665-871(2001).  
 CC -1- SUBCELLULAR LOCATION: Secreted (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE CYSTATIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by, and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL; AL096677; CAC03440.2; -.  
 DR HSSP; P01038; 1A90.  
 DR Genew; HGNC:15958; CRTL.  
 DR InterPro; IPR000010; Cystatin.

DR PROSITE; P500267; CYSTATIN; FALSE NEG.  
 KW Thiol protease inhibitor; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 165  
 FT SITE 93 97  
 FT DISULFID 111 121  
 FT DISULFID 134 154  
 FT CARBOHYD 42 42  
 FT CARBOHYD 54 54  
 FT CARBOHYD 57 57  
 FT CARBOHYD 86 86  
 FT CARBOHYD 114 114  
 FT CARBOHYD 118 118  
 FT CARBOHYD 151 151  
 SQ SEQUENCE 165 AA; 19312 MW; 9D6D685875DAEBA CRC64;  
 Query Match 22.6%; Score 166.5; DB 1; Length 165;  
 Best Local Similarity 26.3%; Pred. No. 1.8e-09;  
 Matches 40; Conservative 34; Mismatches 53; Indels 25; Gaps 4;  
 QY 5 WQALQLLALITLTLMLPYQARKKTPLSVHVAENVAYDSLQWITDQYNKESDDKTHF 64  
 DB 6 WRNPFLLLILALVLSAKLGHFQWEGF--QQLKMSKCN--NNSLTNFFIQSYNNAINDTYLY 62  
 QY 65 RIRVLYKQ-----ROYTDHLEHLVNEQWTTQCKPETTN--CV 102  
 DB 63 RYORLRLRSQWQBERVSHMGLGVHTNSTTDSRLQTTGVGYIVTVXIGWTKCRKRDNTSSSCP 122  
 QY 103 POEBELHKOVNCPFSVPAVPMPEQYKLINKSC 134  
 DB 123 LQSKLKLKSLICSLIYTPWIMYFQLMNNSC 154

Search completed: January 21, 2004, 12:07:20  
 Job time: 18.6774 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: January 21, 2004, 12:25:12 ; Search time 21 Seconds  
(without alignments)  
276.028 Million cell updates/sec

Title: US-09-941-314-2

Perfect score: 137

Sequence: 1 MABPMQALQULLALITLMA.....VFAVPMFEQYKINKSCSSD 137

Scoring table: OLIGO

Searched: 328717 seqs, 42310858 residues

Word size : 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 75 summaries

Database :

Issued Patents AA:\*

1: /cgn2\_6/prodata/1/1aa/5A.COMB.pep:\*

2: /cgn2\_6/prodata/1/1aa/5B.COMB.pep:\*

3: /cgn2\_6/prodata/1/1aa/6A.COMB.pep:\*

4: /cgn2\_6/prodata/1/1aa/6B.COMB.pep:\*

5: /cgn2\_6/prodata/1/1aa/PCBUS.COMB.pep:\*

6: /cgn2\_6/prodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	5.8	432	US-09-252-991A-25540	Sequence 25540, A
2	7	5.1	295	US-09-252-991A-27593	Sequence 27593, A
3	7	5.1	327	US-09-252-991A-29468	Sequence 29468, A
4	7	5.1	415	US-09-006-353A-6	Sequence 6, Appl1
5	7	5.1	415	US-09-573-986-6	Sequence 6, Appl1
6	7	5.1	512	US-09-252-991A-26146	Sequence 26146, A
7	7	5.1	661	US-08-514-014-4	Sequence 4, Appl1
8	7	5.1	661	US-08-833-823-4	Sequence 4, Appl1
9	7	5.1	694	US-09-328-352-7835	Sequence 7835, Ap
10	7	5.1	1844	US-08-851-567B-53	Sequence 53, Appl1
11	7	5.1	2504	US-08-851-567B-12	Sequence 12, Appl1
12	6	4.4	14	US-08-695-692B-12	Sequence 12, Appl1
13	6	4.4	30	US-09-904-196B-7	Sequence 7, Appl1
14	6	4.4	84	US-09-107-532A-5674	Sequence 5674, Ap
15	6	4.4	106	US-08-150-204B-82	Sequence 82, Appl1
16	6	4.4	106	US-08-150-204B-83	Sequence 83, Appl1
17	6	4.4	106	US-08-150-204B-89	Sequence 89, Appl1
18	6	4.4	106	US-08-150-204B-91	Sequence 91, Appl1
19	6	4.4	106	US-08-150-204B-92	Sequence 92, Appl1
20	6	4.4	106	US-08-150-204B-93	Sequence 93, Appl1
21	6	4.4	106	US-08-150-204B-120	Sequence 120, Appl
22	6	4.4	113	US-09-198-452A-450	Sequence 450, Appl
23	6	4.4	126	US-08-513-974B-28	Sequence 28, Appl1
24	6	4.4	126	US-08-776-971-23	Sequence 23, Appl1
25	6	4.4	126	US-09-461-436B-28	Sequence 28, Appl1
26	6	4.4	129	US-09-107-532A-6426	Sequence 6426, Ap
27	6	4.4	132	US-09-252-991A-25272	Sequence 25272, A

## ALIGNMENTS

RESULT 1

US-09-252-991A-25540

Sequence 25540, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252.991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 25540

LENGTH: 422

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-25540

Query Match 5.8%; Score 8; DB 4; Length 422;  
Best Local Similarity 100.0%; Pred. No. 3.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 QALQLLA 13  
Db 147 QALQLLA 154

RESULT 2

US-09-252-991A-27593  
; Sequence 27593, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 27593  
; LENGTH: 295  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27593

Query Match 5.1%; Score 7; DB 4; Length 295;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 65 RIFRVLK 71  
Db 135 RIFRVLK 141

RESULT 3

US-09-252-991A-29468  
; Sequence 29468, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 29468  
; LENGTH: 327  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-29468

Query Match 5.1%; Score 7; DB 4; Length 327;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 LQILLAI 14  
Db 134 LQILLAI 140

RESULT 4

US-09-006-353A-6  
; Sequence 6, Application US/09006353A  
; Patent No. 6261801  
; GENERAL INFORMATION:  
; APPLICANT: WEI, YING-FEI  
; APPLICANT: YU, GUO-LIANG  
; APPLICANT: GENTZ, REINER  
; APPLICANT: RUBEN, STEVEN  
; TITLE OF INVENTION: TUMOR NECROSIS FACTOR RECEPTOR 5  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.  
; STREET: 9410 KEY WEST AVENUE  
; CITY: ROCKVILLE  
; STATE: MD  
; COUNTRY: US  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/006,353A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROOKES, ANDERS A  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PF341  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 415 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-006-353A-6

Query Match 5.1%; Score 7; DB 3; Length 415;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 LLLAILL 16  
Db 222 LLLAILL 228

RESULT 5

US-09-573-986-6  
; Sequence 6, Application US/09573986  
; Patent No. 6455040  
; GENERAL INFORMATION:  
; APPLICANT: Wei, Ying-Fei  
; APPLICANT: Ni, Jian  
; APPLICANT: Gentz, Reiner  
; APPLICANT: Ruben, Steven  
; TITLE OF INVENTION: Tumor Necrosis Factor Receptor 5  
; FILE REFERENCE: 1488.1280004  
; CURRENT APPLICATION NUMBER: US/09/573,986  
; CURRENT FILING DATE: 2000-05-18  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 415  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-573-986-6

Query Match 5.1%; Score 7; DB 4; Length 415;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLALL 16  
DB 222 LLLALL 228

RESULT 6  
US-09-252-991A-26146  
; Sequence 26146, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIORITY FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 26146  
; LENGTH: 512  
; TYPE: PR1  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26146

Query Match 5.1%; Score 7; DB 4; Length 512;  
Best Local Similarity 100.0%; Pred. No. 45;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLLALL 15  
DB 233 QLLALL 239

RESULT 7  
US-08-514-014-4  
; Sequence 4, Application US/08514014  
; Patent No. 5707829  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John  
; APPLICANT: Kelleher, Kerry  
; APPLICANT: Carlin, McKeough  
; TITLE OF INVENTION: DNA SEQUENCES AND SECRETED PROTEINS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc. -- Legal Affairs  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/514,014  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brown, Scott A.  
; REGISTRATION NUMBER: 32,724  
; REFERENCE/DOCKET NUMBER: G16000  
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 661 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-514-014-4

Query Match 5.1%; Score 7; DB 1; Length 661;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLALL 16  
DB 640 LLLALL 646

RESULT 8  
US-08-833-823-4  
; Sequence 4, Application US/08833823  
; Patent No. 5969093  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John  
; APPLICANT: Kelleher, Kerry  
; APPLICANT: Carlin, McKeough  
; TITLE OF INVENTION: DNA SEQUENCES AND SECRETED PROTEINS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc. -- Legal Affairs  
; STREET: 87 CambridgePark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/833,823  
; CLASSIFICATION: 530  
; FILING DATE: 10-APR-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/514,014  
; FILING DATE: 11-AUG-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brown, Scott A.  
; REGISTRATION NUMBER: 32,724  
; REFERENCE/DOCKET NUMBER: G16000  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 876-5851  
; TELEFAX: (617) 876-8224  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 661 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-833-823-4

Query Match 5.1%; Score 7; DB 2; Length 661;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLALL 16  
DB 640 LLLALL 646

RESULT 9  
US-09-328-352-7835  
Sequence 7835, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328.352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 7835  
LENGTH: 694  
TYPE: PRT  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-7835

Query Match 5.1%; Score 7; DB 4; Length 694;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 EGYKILN 131  
Db 106 EGYKILN 112

RESULT 10  
US-08-851-567B-53  
Sequence 53, Application US/08851567B  
Patent No. 6528484  
GENERAL INFORMATION:  
APPLICANT: Ensign, Jerald C  
APPLICANT: Bowen, David J  
APPLICANT: Petell, James  
APPLICANT: Fatig, Raymond  
APPLICANT: Schoonover, Sue  
APPLICANT: French-Constant, Richard  
APPLICANT: Rocheleau, Thomas A.  
APPLICANT: Blackburn, Michael B.  
APPLICANT: Hey, Timothy D.  
APPLICANT: Merlo, Donald J.  
APPLICANT: Orr, Gregory L.  
APPLICANT: Roberts, Jean L.  
APPLICANT: Strickland, James A.  
APPLICANT: Guo, Lining  
APPLICANT: Cliche, Todd A.  
APPLICANT: Sukhapiinda, Kiti Sri  
TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus  
NUMBER OF SEQUENCES: 88  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dow Agrosciences Patent Department  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: IN  
COUNTRY: US  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/851.567B  
FILING DATE: 05-MAY-1997  
CLASSIFICATION: 514  
APPLICATION NUMBER: US 08/063.615  
FILING DATE: 18-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/395.497  
FILING DATE: 28-FEB-1995

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/007,255  
FILING DATE: 06-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/608,423  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/705,484  
FILING DATE: 28-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960296.93804  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 608-251-5000  
TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1844 amino acids  
TYPE: amino acids  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-851-567B-53

Query Match 5.1%; Score 7; DB 4; Length 1844;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 LITLML 21  
Db 712 LITLML 718

RESULT 11  
US-08-851-567B-12  
Sequence 12, Application US/08851567B  
Patent No. 6528484  
GENERAL INFORMATION:  
APPLICANT: Ensign, Jerald C  
APPLICANT: Bowen, David J  
APPLICANT: Petell, James  
APPLICANT: Fatig, Raymond  
APPLICANT: Schoonover, Sue  
APPLICANT: French-Constant, Richard  
APPLICANT: Rocheleau, Thomas A.  
APPLICANT: Blackburn, Michael B.  
APPLICANT: Hey, Timothy D.  
APPLICANT: Merlo, Donald J.  
APPLICANT: Orr, Gregory L.  
APPLICANT: Roberts, Jean L.  
APPLICANT: Strickland, James A.  
APPLICANT: Guo, Lining  
APPLICANT: Cliche, Todd A.  
APPLICANT: Sukhapiinda, Kiti Sri  
TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus  
NUMBER OF SEQUENCES: 88  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dow Agrosciences Patent Department  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: IN  
COUNTRY: US  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/851.567B  
FILING DATE: 05-MAY-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/063,615  
FILING DATE: 18-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/395,497  
FILING DATE: 28-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/007,255  
FILING DATE: 06-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/608,423  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/705,484  
FILING DATE: 28-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J  
REGISTRATION NUMBER: 27386  
TELEPHONE/DOCKET NUMBER: 960296.93804  
TELEPHONE: 608-251-5000  
TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2504 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-851-567B-12

Query Match 5.1%; Score 7; DB 4; Length 2504;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 LITLML 21  
DB 799 LITLML 805

RESULT 12  
US-08-695-692B-12  
Sequence 12, Application US/08695692B  
Patent No. 651498  
GENERAL INFORMATION:  
APPLICANT: Per Antonsson, Per Bjork, Mikael Dohlsten,  
APPLICANT: Johan Hansson, Terje Kalland, Lars  
APPLICANT: Abrahamson and Goran Forsberg  
TITLE OF INVENTION: MODIFIED/CHIMERIC SUPRANTIGENS  
TITLE OF INVENTION: AND THEIR USE  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pravel, Hewitt, Kimball & Krieger  
STREET: 1177 West Loop South, 10th Floor  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77027-9095  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/695,692B  
FILING DATE: August 12, 1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9601245-5  
FILING DATE: March 29, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Krieger, Paul E.  
REGISTRATION NUMBER: 25,866  
REFERENCE/DOCKET NUMBER: 41986/1  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 713-850-0909  
TELEFAX: 713-850-0165  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-695-692B-12

Query Match 4.4%; Score 6; DB 4; Length 14;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 NKESDD 60  
DB 4 NKESDD 9

RESULT 13  
US-09-904-196B-7  
Sequence 7, Application US/09904196B  
Patent No. 6555660  
GENERAL INFORMATION:  
APPLICANT: NISSEN, TORBEN LAUSGAARD  
APPLICANT: ANDERSEN, KIM VILBOUR  
APPLICANT: HANSEN, CHRISTIAN KARSTEN  
APPLICANT: MIKELSEN, JAN MOLLER  
TITLE OF INVENTION: G-CSF CONJUGATES  
FILE REFERENCE: 31-000700US  
CURRENT APPLICATION NUMBER: US/09/904,196B  
CURRENT FILING DATE: 2001-07-11  
PRIOR APPLICATION NUMBER: US/09/760,008  
PRIOR FILING DATE: 2001-05-11  
PRIOR APPLICATION NUMBER: 60/176,376  
PRIOR FILING DATE: 2000-01-14  
PRIOR APPLICATION NUMBER: 60/189,506  
PRIOR FILING DATE: 2000-03-15  
PRIOR APPLICATION NUMBER: 60/215,644  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DK PA 2000 00024  
PRIOR FILING DATE: 2000-01-10  
PRIOR APPLICATION NUMBER: DK PA 2000 00341  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: DK PA 2000 00943  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 7  
LENGTH: 30  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-904-196B-7

Query Match 4.4%; Score 6; DB 4; Length 30;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ALQULL 12  
DB 14 ALQULL 19

RESULT 14  
US-09-107-532A-5674  
Sequence 5674, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107.532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneka  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 5674:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 84 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...84  
SEQUENCE DESCRIPTION: SEQ ID NO: 5674:  
US-09-107-532A-5674

Query Match 4.4%; Score 6; DB 4; Length 84;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 LILTIM 19  
DB 35 LILTIM 40

RESULT 15  
US-08-150-204E-82  
Sequence 82, Application US/08150204E  
Patent No. 6538126  
GENERAL INFORMATION:  
APPLICANT: CHO, Joong Myung  
LEE, Yong Beom  
PARK, Young Moo  
LIM, Kook Jin  
CHOI, Deog Young  
SO, Hong Seob  
KIM, Chun Hyung  
KIM, Sung Taek  
YANG, Jae Young  
TITLE OF INVENTION: HEPATITIS C DIAGNOSTICS AND VACCINES  
NUMBER OF SEQUENCES: 128  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: YANG, Jae Young  
STREET: 386-1, Doryong-dong, Yuseong-gu  
CITY: Daejeon  
STATE: Daejeon  
COUNTRY: Republic of Korea  
ZIP: 305-340

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5inch, 1.44MB storage  
COMPUTER: IBM PC/pendium  
OPERATING SYSTEM: windows  
SOFTWARE: Microsoft word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/150,204E  
FILING DATE: 20-Apr-1994  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: KR 91-9510  
FILING DATE: 10-JUN-1991  
APPLICATION NUMBER: KR 91-13601  
FILING DATE: 6-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Shahan Islam, Esq.  
REGISTRATION NUMBER: 32,507  
REFERENCE/DOCKET NUMBER: 2695/FLK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 940-8564  
TELEFAX: (212) 940-8776  
INFORMATION FOR SEQ ID NO: 82  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 82  
US-08-150-204E-82

Query Match 4.4%; Score 6; DB 4; Length 106;  
Best Local Similarity 100.0%; Pred. No. 1,1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LILAIL 15  
DB 64 LILAIL 69

RESULT 16  
US-08-150-204E-83  
Sequence 83, Application US/08150204E  
Patent No. 6538126  
GENERAL INFORMATION:  
APPLICANT: CHO, Joong Myung  
LEE, Yong Beom  
PARK, Young Moo  
LIM, Kook Jin  
CHOI, Deog Young  
SO, Hong Seob  
KIM, Chun Hyung  
KIM, Sung Taek  
YANG, Jae Young  
TITLE OF INVENTION: HEPATITIS C DIAGNOSTICS AND VACCINES  
NUMBER OF SEQUENCES: 128  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: YANG, Jae Young  
STREET: 386-1, Doryong-dong, Yuseong-gu  
CITY: Daejeon  
STATE: Daejeon  
COUNTRY: Republic of Korea  
ZIP: 305-340  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5inch, 1.44MB storage  
COMPUTER: IBM PC/pendium  
OPERATING SYSTEM: windows  
SOFTWARE: Microsoft word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/150,204E  
FILING DATE: 20-Apr-1994  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION NUMBER: KR 91-9510

FILING DATE: 10-JUN-1991  
APPLICATION NUMBER: KR 91-13601  
FILING DATE: 6-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Shahan Islam, Esq.  
REGISTRATION NUMBER: 32,507  
REFERENCE/DOCKET NUMBER: 2695/FLK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 940-8564  
TELEFAX: (212) 940-8776  
INFORMATION FOR SEQ ID NO: 83  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 83  
US-08-150-204E-83

Query Match 4.4%; Score 6; DB 4; Length 106;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 L1LAIL 15  
Db 64 L1LAIL 69

RESULT 17  
US-08-150-204E-89  
Sequence 89, Application US/08150204E  
Patent No. 6538126  
GENERAL INFORMATION:  
APPLICANT: CHO, Joong Myung  
LEE, Yong Beom  
PARK, Young Moo  
LIM, Kook Jin  
CHOI, Deog Young  
SO, Hong Seob  
KIM, Chun Hyung  
KIM, Sung Taek  
YANG, Jae Young  
TITLE OF INVENTION: HEPATITIS C DIAGNOSTICS AND VACCINES  
NUMBER OF SEQUENCES: 128  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: YANG, Jae Young  
STREET: 386-1, Doryong-dong, Yuseong-gu  
CITY: Daejeon  
STATE: Daejeon  
COUNTRY: Republic of Korea  
ZIP: 305-340  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5inch, 1.44MB storage  
COMPUTER: IBM PC/pentium  
OPERATING SYSTEM: Windows  
SOFTWARE: Microsoft Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/150, 204E  
FILING DATE: 20-Apr-1994  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: KR 91-9510  
FILING DATE: 10-JUN-1991  
APPLICATION NUMBER: KR 91-13601  
FILING DATE: 6-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Shahan Islam, Esq.  
REGISTRATION NUMBER: 32,507  
REFERENCE/DOCKET NUMBER: 2695/FLK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 940-8564  
TELEFAX: (212) 940-8776  
INFORMATION FOR SEQ ID NO: 89

SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 89  
US-08-150-204E-89

Query Match 4.4%; Score 6; DB 4; Length 106;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 L1LAIL 15  
Db 64 L1LAIL 69

RESULT 18  
US-08-150-204E-91  
Sequence 91, Application US/08150204E  
Patent No. 6538126  
GENERAL INFORMATION:  
APPLICANT: CHO, Joong Myung  
LEE, Yong Beom  
PARK, Young Moo  
LIM, Kook Jin  
CHOI, Deog Young  
SO, Hong Seob  
KIM, Chun Hyung  
KIM, Sung Taek  
YANG, Jae Young  
TITLE OF INVENTION: HEPATITIS C DIAGNOSTICS AND VACCINES  
NUMBER OF SEQUENCES: 128  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: YANG, Jae Young  
STREET: 386-1, Doryong-dong, Yuseong-gu  
CITY: Daejeon  
STATE: Daejeon  
COUNTRY: Republic of Korea  
ZIP: 305-340  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5inch, 1.44MB storage  
COMPUTER: IBM PC/pentium  
OPERATING SYSTEM: Windows  
SOFTWARE: Microsoft Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/150, 204E  
FILING DATE: 20-Apr-1994  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: KR 91-9510  
FILING DATE: 10-JUN-1991  
APPLICATION NUMBER: KR 91-13601  
FILING DATE: 6-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Shahan Islam, Esq.  
REGISTRATION NUMBER: 32,507  
REFERENCE/DOCKET NUMBER: 2695/FLK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 940-8564  
TELEFAX: (212) 940-8776  
INFORMATION FOR SEQ ID NO: 91  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 91  
US-08-150-204E-91

Query Match 4.4%; Score 6; DB 4; Length 106;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIL 15  
|||||  
Db 64 LLLAIL 69

## RESULT 19

US-08-150-204E-92  
; Sequence 92, Application US/08150204E  
; Patent No. 6538126

## GENERAL INFORMATION:

APPLICANT: CHO, Joong Myung

LEE, Yong Beom

PARK, Young Moo

LIM, Kook Jin

CHOI, Deog Young

SO, Hong Seob

KIM, Chun Hyung

KIM, Sung Taek

YANG, Jae Young

TITLE OF INVENTION: HEPATITIS C DIAGNOSTICS AND VACCINES

NUMBER OF SEQUENCES: 128

CORRESPONDENCE ADDRESS:

ADDRESSEE: YANG, Jae Young

STREET: 386-1, Doryong-dong, Yuseong-gu

CITY: Daejeon

STATE: Daejeon

COUNTRY: Republic of Korea

ZIP: 305-340

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5inch, 1.44MB storage

COMPUTER: IBM PC/pentium

OPERATING SYSTEM: Windows

SOFTWARE: Microsoft word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/150,204E

FILING DATE: 20-Apr-1994

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: KR 91-9510

FILING DATE: 10-JUN-1991

APPLICATION NUMBER: KR 91-13601

FILING DATE: 6-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: Shahan Islam, Esq.

REGISTRATION NUMBER: 32,507

REFERENCE/DOCKET NUMBER: 2695/FLK

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 940-8564

TELEFAX: (212) 940-8776

INFORMATION FOR SEQ ID NO: 92

SEQUENCE CHARACTERISTICS:

LENGTH: 106 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULAR TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 92

US-08-150-204E-92

QY 10 LLLAIL 15  
|||||  
Db 64 LLLAIL 69

## RESULT 20

US-08-150-204E-93  
; Sequence 93, Application US/08150204E  
; Patent No. 6538126

## GENERAL INFORMATION:

Query Match 4.4%; Score 6; DB 4; Length 106;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

APPLICANT: CHO, Joong Myung

LEE, Yong Beom

PARK, Young Moo

LIM, Kook Jin

CHOI, Deog Young

SO, Hong Seob

KIM, Chun Hyung

KIM, Sung Taek

YANG, Jae Young

TITLE OF INVENTION: HEPATITIS C DIAGNOSTICS AND VACCINES

NUMBER OF SEQUENCES: 128

CORRESPONDENCE ADDRESS:

ADDRESSEE: YANG, Jae Young

STREET: 386-1, Doryong-dong, Yuseong-gu

CITY: Daejeon

STATE: Daejeon

COUNTRY: Republic of Korea

ZIP: 305-340

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5inch, 1.44MB storage

COMPUTER: IBM PC/pentium

OPERATING SYSTEM: Windows

SOFTWARE: Microsoft word

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/150,204E

FILING DATE: 20-Apr-1994

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: KR 91-9510

FILING DATE: 10-JUN-1991

APPLICATION NUMBER: KR 91-13601

FILING DATE: 6-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: Shahan Islam, Esq.

REGISTRATION NUMBER: 32,507

REFERENCE/DOCKET NUMBER: 2695/FLK

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 940-8564

TELEFAX: (212) 940-8776

INFORMATION FOR SEQ ID NO: 93

SEQUENCE CHARACTERISTICS:

LENGTH: 106 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULAR TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 93

US-08-150-204E-93

QY 10 LLLAIL 15  
|||||  
Db 64 LLLAIL 69

## RESULT 21

US-08-150-204E-120  
; Sequence 120, Application US/08150204E  
; Patent No. 6538126

## GENERAL INFORMATION:

APPLICANT: CHO, Joong Myung

LEE, Yong Beom

PARK, Young Moo

LIM, Kook Jin

CHOI, Deog Young

SO, Hong Seob

KIM, Chun Hyung

KIM, Sung Taek

YANG, Jae Young

TITLE OF INVENTION: HEPATITIS C DIAGNOSTICS AND VACCINES

NUMBER OF SEQUENCES: 128

CORRESPONDENCE ADDRESS:  
ADDRESSEE: YANG, Jae Young  
STREET: 386-1, Doryong-dong, Yuseong-gu  
CITY: Daejeon  
STATE: Daejeon  
COUNTRY: Republic of Korea  
ZIP: 305-340  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5inch, 1.44MB storage  
COMPUTER: IBM PC/Pentium  
OPERATING SYSTEM: Windows  
SOFTWARE: Microsoft Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/150, 204B  
FILING DATE: 20-Apr-1994  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: KR 91-9510  
FILING DATE: 10-JUN-1991  
APPLICATION NUMBER: KR 91-13601  
FILING DATE: 6-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Shahan Ielam, Esq.  
REGISTRATION NUMBER: 32,507  
REFERENCE/DOCKET NUMBER: 2695/FLK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 940-8564  
TELEFAX: (212) 940-8776  
INFORMATION FOR SEQ ID NO: 120  
SEQUENCE CHARACTERISTICS:  
LENGTH: 106 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
OTHER INFORMATION: JHCY-NCI  
SEQUENCE DESCRIPTION: SEQ ID NO: 120  
US-08-150-204B-120

Query Match 4.4%; Score 6; DB 4; Length 106;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLALL 15  
Db 64 LLLALL 69

RESULT 22  
US-09-198-452A-450  
Sequence 450, Application US/09198452A  
Patent No. 6559294  
GENERAL INFORMATION:  
APPLICANT: Griffiths, R.  
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention  
FILE REFERENCE: 9710-003-999  
CURRENT APPLICATION NUMBER: US/09/198, 452A  
FILING DATE: 1998-11-24  
NUMBER OF SEQ ID NOS: 6849  
SEQ ID NO 450  
LENGTH: 113  
TYPE: PRT  
ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-450

Query Match 4.4%; Score 6; DB 4; Length 113;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LALLLT 17  
|||||

Db 46 LALLLT 51

RESULT 23  
US-08-513-974B-28  
Sequence 28, Application US/08513974B  
Patent No. 6114139  
GENERAL INFORMATION:  
APPLICANT: Hinuma, Shuji  
APPLICANT: Hosoya, Masaki  
APPLICANT: Fujii, Ryo  
APPLICANT: Ohtsuki, Tetsuya  
APPLICANT: Fukusumi, Shoji  
APPLICANT: Ohgi, Kazuhiro  
TITLE OF INVENTION: G PROTEIN COUPLED RECEPTOR PROTEIN,  
TITLE OF INVENTION: PRODUCTION, AND USE THEREOF  
NUMBER OF SEQUENCES: 380  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP  
STREET: 130 Water Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/513, 974B  
FILING DATE: 14-SEP-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP95/01599  
FILING DATE: 10-AUG-1995  
APPLICATION NUMBER: JP 7-093989  
FILING DATE: 19-AUG-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7-057186  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7-007177  
FILING DATE: 20-JAN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-326611  
FILING DATE: 28-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-270017  
FILING DATE: 02-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-236357  
FILING DATE: 30-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-236356  
FILING DATE: 30-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-189274  
FILING DATE: 11-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-189273  
FILING DATE: 11-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-189272  
FILING DATE: 11-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Resnick, David S.  
REGISTRATION NUMBER: 34,235  
REFERENCE/DOCKET NUMBER: 45753  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 126 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-513-974B-28

Query Match 4.4%; Score 6; DB 3; Length 126;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 L1A1L1 16  
Db 78 L1A1L1 83

## RESULT 24

US-08-776-971-23  
Sequence 23, Application US/0876971B  
Patent No. 6228984

## GENERAL INFORMATION:

APPLICANT: Hinuma, Shuji

Habata, Yugo

Kawamata, Yuji

Hosoya, Masaki

Fujii, Ryo

Fukushima, Shoichi

Kitada, Chieko

TITLE OF INVENTION: POLYPEPTIDES, THEIR PRODUCTION AND USE

NUMBER OF SEQUENCES: 140

CORRESPONDENCE ADDRESSES:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP

STREET: 130 Water Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/776,971B

FILING DATE: 06-Feb-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP96/03821

FILING DATE: 28-DEC-1996

APPLICATION NUMBER: JP 7/343371

FILING DATE: 28-DEC-1995

APPLICATION NUMBER: JP 8/59419

FILING DATE: 15-MAR-1996

APPLICATION NUMBER: JP 8/211805

FILING DATE: 12-AUG-1996

APPLICATION NUMBER: JP 8/246573

FILING DATE: 18-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.

REGISTRATION NUMBER: 27,026

REFERENCE/DOCKET NUMBER: 47176

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 126 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal  
SEQUENCE DESCRIPTION: SEQ ID NO: 23:  
US-08-776-971-23

Query Match 4.4%; Score 6; DB 3; Length 126;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 L1A1L1 16  
Db 78 L1A1L1 83

## RESULT 25

US-09-461-436B-28  
Sequence 28, Application US/09461436B  
Patent No. 6538107

GENERAL INFORMATION:

APPLICANT: Shuji Hinuma

Yasuaki Ito

Ryo Fujii

TITLE OF INVENTION: G Protein Coupled Receptor Protein,

Production, And Use Thereof

NUMBER OF SEQUENCES: 61

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Edwards & Angell, LLP

STREET: 101 Federal Street

CITY: BOSTON

STATE: MA

COUNTRY: USA

ZIP: 02209

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/461,436B

FILING DATE: 14-Dec-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/513,974

FILING DATE: 14-SEP-1995

APPLICATION NUMBER: PCT/JP95/01599

FILING DATE: 10-AUG-1995

APPLICATION NUMBER: 7-093989

FILING DATE: 19-APR-1995

APPLICATION NUMBER: 7-057186

FILING DATE: 16-MAR-1995

APPLICATION NUMBER: 7-007177

FILING DATE: 20-JAN-1995

APPLICATION NUMBER: 6-326611

FILING DATE: 28-DEC-1994

APPLICATION NUMBER: 6-270017

FILING DATE: 02-NOV-1994

APPLICATION NUMBER: 6-236357

FILING DATE: 30-SEP-1994

APPLICATION NUMBER: 6-236356

FILING DATE: 30-SEP-1994

APPLICATION NUMBER: 6-189274

FILING DATE: 11-AUG-1994

APPLICATION NUMBER: 6-189273

FILING DATE: 11-AUG-1994

APPLICATION NUMBER: 6-189272

FILING DATE: 11-AUG-1994

ATTORNEY/AGENT INFORMATION:

NAME: CONLIN, DAVID G.

REGISTRATION NUMBER: <Unknown>

REFERENCE/DOCKET NUMBER: 45753 DIV2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-439-4444

TELEFAX: 617-439-4170

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:  
LENGTH: 126 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 28;  
US-09-461-436B-28

Query Match 4.4%; Score 6; DB 4; Length 126;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 LLA1LL 16  
|||  
Db 78 LLA1LL 83

RESULT 26  
US-09-107-532A-6426  
Sequence 6426, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucet-Re-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Arinello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 6426:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 129 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...129  
SEQUENCE DESCRIPTION: SEQ ID NO: 6426;  
US-09-107-532A-6426

Query Match 4.4%; Score 6; DB 4; Length 129;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 ESDDKY 62  
|||  
Db 21 ESDDKY 26

RESULT 27  
US-09-252-991A-25272  
Sequence 25272, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 25272  
LENGTH: 132  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-25272

Query Match 4.4%; Score 6; DB 4; Length 132;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 LPYQAR 26  
|||  
Db 71 LPYQAR 76

RESULT 28  
US-08-513-974B-366  
Sequence 366, Application US/08513974B  
Patent No. 6114139  
GENERAL INFORMATION:  
APPLICANT: Hinuma, Shuji  
APPLICANT: Hosoya, Masaki  
APPLICANT: Fujii, Ryo  
APPLICANT: Ohtaki, Tetsuya  
APPLICANT: Fukusumi, Shoji  
APPLICANT: Ohgi, Kazuhito  
TITLE OF INVENTION: G PROTEIN COUPLED RECEPTOR PROTEIN,  
PRODUCTION, AND USE THEREOF  
NUMBER OF SEQUENCES: 380  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP  
STREET: 130 Water Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: IBM PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/513,974B  
FILING DATE: 14-SEP-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP95/01599  
FILING DATE: 10-AUG-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7-093989  
FILING DATE: 19-AUG-1995  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 7-057186  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7-007177  
FILING DATE: 20-JAN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-326611  
FILING DATE: 28-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-270017  
FILING DATE: 02-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-236357  
FILING DATE: 30-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-236356  
FILING DATE: 30-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-189274  
FILING DATE: 11-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-189273  
FILING DATE: 11-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-189272  
FILING DATE: 11-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Resnick, David S.  
REGISTRATION NUMBER: 34,235  
REFERENCE/DOCKET NUMBER: 45753  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
INFORMATION FOR SEQ ID NO: 366:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 144 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-513-974B-366

Query Match 4.4%; Score 6; DB 3; Length 144;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 LLA1LL 16  
DB 87 LLA1LL 92

RESULT 29  
US-08-513-974B-369  
Sequence 369, Application US/08513974B  
Patent No. 6114139  
GENERAL INFORMATION:  
APPLICANT: Hinuma, Shuji  
APPLICANT: Hosoya, Masaki  
APPLICANT: Fujii, Ryo  
APPLICANT: Ohtaki, Tetsuya  
APPLICANT: Fukusumi, Shoji  
APPLICANT: Ohgi, Kazuhiko  
TITLE OF INVENTION: G PROTEIN COUPLED RECEPTOR PROTEIN,  
TITLE OF INVENTION: PRODUCTION, AND USE THEREOF  
NUMBER OF SEQUENCES: 380  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP  
STREET: 130 Water Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/513,974B  
FILING DATE: 14-SEP-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP95/01599  
FILING DATE: 10-AUG-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7-093989  
FILING DATE: 19-AUG-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7-057186  
FILING DATE: 16-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 7-007177  
FILING DATE: 20-JAN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-326611  
FILING DATE: 28-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-270017  
FILING DATE: 02-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-236357  
FILING DATE: 30-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-236356  
FILING DATE: 30-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-189274  
FILING DATE: 11-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-189273  
FILING DATE: 11-AUG-1945  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 6-189272  
FILING DATE: 11-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Resnick, David S.  
REGISTRATION NUMBER: 34,235  
REFERENCE/DOCKET NUMBER: 45753  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
INFORMATION FOR SEQ ID NO: 369:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 144 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-513-974B-369

Query Match 4.4%; Score 6; DB 3; Length 144;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 LLA1LL 16  
DB 87 LLA1LL 92

RESULT 30  
US-08-776-971-106  
Sequence 106, Application US/08776971B  
Patent No. 6228984  
GENERAL INFORMATION:  
APPLICANT: Hinuma, Shuji  
APPLICANT: Habata, Yugo  
APPLICANT: Kawamata, Yuji

Hosoya, Masaki  
Fuji, Ryo  
Fukusumi, Shoji  
Kitada, Chieko  
TITLE OF INVENTION: POLYPROTEINS, THEIR PRODUCTION AND USE  
NUMBER OF SEQUENCES: 140  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP  
STREET: 130 Water Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/776,971B  
FILING DATE: 06-Feb-1997  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP96/03821  
FILING DATE: 28-DEC-1996  
APPLICATION NUMBER: JP 7/343371  
FILING DATE: 28-DEC-1995  
APPLICATION NUMBER: JP 8/59419  
FILING DATE: 15-MAR-1996  
APPLICATION NUMBER: JP 8/211805  
FILING DATE: 12-AUG-1996  
APPLICATION NUMBER: JP 8/246573  
FILING DATE: 18-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Conlin, David G.  
REGISTRATION NUMBER: 27,026  
REFERENCE/DOCKET NUMBER: 47176  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
INFORMATION FOR SEQ ID NO: 106:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 144 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
SEQUENCE DESCRIPTION: SEQ ID NO: 106:  
US-08-776-971-106  
Query Match 4.4%; Score 6; DB 3; Length 144;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 11 LLAILL 16  
DB 87 LLAILL 92  
RESULT 31  
US-08-776-971-109  
Sequence 109, Application US/08776971B  
Patent No. 6228984  
GENERAL INFORMATION:  
APPLICANT: Hinuma, Shuji  
Hinata, Yugo  
Kawamata, Yuji  
Hosoya, Masaki  
Fuji, Ryo  
Fukusumi, Shoji  
Kitada, Chieko  
TITLE OF INVENTION: POLYPROTEINS, THEIR PRODUCTION AND USE

NUMBER OF SEQUENCES: 140  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP  
STREET: 130 Water Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/776,971B  
FILING DATE: 06-Feb-1997  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP96/03821  
FILING DATE: 28-DEC-1996  
APPLICATION NUMBER: JP 7/343371  
FILING DATE: 28-DEC-1995  
APPLICATION NUMBER: JP 8/59419  
FILING DATE: 15-MAR-1996  
APPLICATION NUMBER: JP 8/211805  
FILING DATE: 12-AUG-1996  
APPLICATION NUMBER: JP 8/246573  
FILING DATE: 18-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Conlin, David G.  
REGISTRATION NUMBER: 27,026  
REFERENCE/DOCKET NUMBER: 47176  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
INFORMATION FOR SEQ ID NO: 109:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 144 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 109:  
US-08-776-971-109  
Query Match 4.4%; Score 6; DB 3; Length 144;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 11 LLAILL 16  
DB 87 LLAILL 92  
RESULT 32  
US-08-832-535-11  
Sequence 11, Application US/08832535  
Patent No. 5919658  
GENERAL INFORMATION:  
APPLICANT: NI, JIAN  
LI, HAODONG  
APPLICANT: YU, GUO-LIANG  
APPLICANT: GENTZ, REINER L  
TITLE OF INVENTION: HUMAN CYSTATIN P  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HUMAN GENOME SCIENCES, INC.  
STREET: 9410 KEY WEST AVENUE  
CITY: ROCKVILLE  
STATE: MD  
COUNTRY: US  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/832,535  
FILING DATE: 03-APR-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: KIMBALL, PAUL C.  
REGISTRATION NUMBER: 34,610  
REFERENCE/DOCKET NUMBER: PF265  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (201) 994-1700  
TELEFAX: (201) 994-1744  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 145 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-832-535-11

Query Match 4.4%; Score 6; DB 2; Length 145;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIL 15  
DB 10 LLLAIL 15

RESULT 33  
US-08-791-522-3  
Sequence 3, Application US/08791522  
Patent No. 5935817  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goll, Surya K.  
TITLE OF INVENTION: NOVEL HUMAN CYSTATIN-LIKE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/791,522  
FILING DATE: Filed Herewith  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0193 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 146 amino acids  
TYPE: amino acid  
STRANDEDNESS: single

TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: Genbank  
CLONE: 101387  
US-08-791-522-3

Query Match 4.4%; Score 6; DB 2; Length 146;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIL 15  
DB 10 LLLAIL 15

RESULT 34  
US-08-744-138-3  
Sequence 3, Application US/08744138  
Patent No. 6011012  
GENERAL INFORMATION:  
APPLICANT: Gentz, Reiner L.  
APPLICANT: Ni, Jian  
APPLICANT: Rosen, Craig A.  
APPLICANT: Yu, Guo-Liang  
TITLE OF INVENTION: Human Cystatin E  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/744,138  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PF202P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301 309 8512  
TELEFAX: 301 309 8504  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 146 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Cystatin C  
US-08-744-138-3

Query Match 4.4%; Score 6; DB 3; Length 146;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIL 15  
DB 10 LLLAIL 15

RESULT 35  
US-09-019-485-4  
Sequence 4, Application US/09019485

Patent No. 6066617  
GENERAL INFORMATION:  
APPLICANT: Li, Haodong  
APPLICANT: Yu, Guo-liang  
APPLICANT: Gentz, Reiner  
APPLICANT: Ni, Jian  
TITLE OF INVENTION: Cystatin F  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: MD  
COUNTRY: US  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/019,485  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Benson, Robert H.  
REGISTRATION NUMBER: 30,446  
REFERENCE/DOCKET NUMBER: PF265P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 3013098504  
TELEFAX: 3013098439  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 146 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-019-485-4

Query Match 4.4%; Score 6; DB 3; Length 146;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIL 15  
DB 10 LLLAIL 15

RESULT 36  
US-09-314-777-3  
Sequence 3, Application US/09314777  
Patent No. 6110686  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goll, Surya K.  
TITLE OF INVENTION: NOVEL HUMAN CYSTATIN-LIKE  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/314,777

FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/791,522  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0193 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 146 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 181387  
US-09-314-777-3

Query Match 4.4%; Score 6; DB 3; Length 146;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIL 15  
DB 10 LLLAIL 15

RESULT 37  
US-09-431-480-6  
Sequence 6, Application US/09431480  
Patent No. 6235708  
GENERAL INFORMATION:  
APPLICANT: Holloway, James L.  
APPLICANT: Feldhaus, Andrew  
TITLE OF INVENTION: TESTIS SPECIFIC CYSTATIN-LIKE PROTEIN CYSTATIN T  
FILE REFERENCE: 98-72  
CURRENT APPLICATION NUMBER: US/09/431,480  
CURRENT FILING DATE: 1999-11-01  
EARLIER APPLICATION NUMBER: 60/109,217  
EARLIER FILING DATE: 1998-11-20  
EARLIER APPLICATION NUMBER: 60/156,382  
EARLIER FILING DATE: 1999-09-28  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 6  
LENGTH: 146  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-431-480-6

Query Match 4.4%; Score 6; DB 3; Length 146;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIL 15  
DB 10 LLLAIL 15

RESULT 38  
US-09-617-302-6  
Sequence 6, Application US/09617302  
Patent No. 6245529  
GENERAL INFORMATION:  
APPLICANT: Holloway, James L.  
APPLICANT: Feldhaus, Andrew  
TITLE OF INVENTION: TESTIS SPECIFIC CYSTATIN-LIKE PROTEIN CYSTATIN T  
FILE REFERENCE: 98-72 C1

CURRENT APPLICATION NUMBER: US/09/617,302  
CURRENT FILING DATE: 2000-07-17  
PRIOR APPLICATION NUMBER: 09/431,480  
PRIOR FILING DATE: 1999-11-01  
PRIOR APPLICATION NUMBER: 60/109,217  
PRIOR FILING DATE: 1998-11-20  
PRIOR APPLICATION NUMBER: 60/156,382  
PRIOR FILING DATE: 1999-09-28  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 6  
LENGTH: 146  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-617-302-6

Query Match 4.4%; Score 6; DB 3; Length 146;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIL 15  
DB 10 LLLAIL 15

RESULT 39  
US-09-241-376-3  
Sequence 3, Application US/09241376  
Patent No. 6300477  
GENERAL INFORMATION:  
APPLICANT: Gentz, Reiner L.  
APPLICANT: Ni, Jian  
APPLICANT: Rosen, Craig A.  
APPLICANT: Yu, Guo-Liang  
TITLE OF INVENTION: Human Cystatin B  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/241,376  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/744,138  
FILING DATE: 05-NOV-1996  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/461,030  
FILING DATE: 05-JUN-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PP202PID1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301 309 8504  
TELEFAX: 301 309 8512  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 146 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear

MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: Cystatin C  
US-09-241-376-3

Query Match 4.4%; Score 6; DB 4; Length 146;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIL 15  
DB 10 LLLAIL 15

RESULT 40  
US-09-528-436B-3  
Sequence 3, Application US/09528436B  
Patent No. 6576745  
GENERAL INFORMATION:  
APPLICANT: Li, et al.  
TITLE OF INVENTION: Human Cystatin F  
FILE REFERENCE: PP265PID1  
CURRENT APPLICATION NUMBER: US/09/528,436B  
CURRENT FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: 09/019,485  
PRIOR FILING DATE: 1998-01-29  
PRIOR APPLICATION NUMBER: 08/832,535  
PRIOR FILING DATE: 1999-04-03  
PRIOR APPLICATION NUMBER: 60/014,795  
PRIOR FILING DATE: 1996-04-03  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: Patentin version 3.2  
SEQ ID NO 3  
LENGTH: 146  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-528-436B-3

Query Match 4.4%; Score 6; DB 4; Length 146;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIL 15  
DB 10 LLLAIL 15

RESULT 41  
US-09-886-319A-47  
Sequence 47, Application US/09886319A  
Patent No. 6586185  
GENERAL INFORMATION:  
APPLICANT: Wolf, Eckard  
APPLICANT: Werner, Sabine  
APPLICANT: Halle, Jörn-Peter  
APPLICANT: Regenbogen, Johannes  
APPLICANT: Goppelt, Andreas  
TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for  
TITLE OF INVENTION: the Diagnosis or Treatment of Skin Disorders and Wound  
TITLE OF INVENTION: Healing and for the Identification of Pharmacologically  
TITLE OF INVENTION: Active Substances  
FILE REFERENCE: 50125/014002  
CURRENT APPLICATION NUMBER: US/09/886,319A  
CURRENT FILING DATE: 2001-06-20  
PRIOR APPLICATION NUMBER: US 60/222,081  
PRIOR FILING DATE: 2000-08-01  
PRIOR APPLICATION NUMBER: DE 10030149.5  
PRIOR FILING DATE: 2000-06-20  
NUMBER OF SEQ ID NOS: 84  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 47  
LENGTH: 146

TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-886-319A-47

Query Match  
Best Local Similarity 100.0%; Score 6; DB 4; Length 146;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIL 15  
Db 10 LLLAIL 15

RESULT 42  
PCT-US95-07135-9

Sequence 9, Application PC/TUS9507135  
GENERAL INFORMATION:  
APPLICANT: Ni, Jian  
APPLICANT: Yu, Guo-liang  
APPLICANT: Gentz, Reiner  
APPLICANT: Rosen, Craig  
TITLE OF INVENTION: HUMAN CYSTATIN B  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
ADDRESSEE: CECCHI, STUART & OLSTEIN  
STREET: 6 Becker Farm Road  
CITY: Roseland  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/07135  
FILING DATE: 05-JUN-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: MULLINS, J.G.  
REGISTRATION NUMBER: 33,073  
REFERENCE/DOCKET NUMBER: 325800-400  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 146 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-07135-9

Query Match  
Best Local Similarity 100.0%; Score 6; DB 5; Length 146;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIL 15  
Db 10 LLLAIL 15

RESULT 43  
5432264-6  
Patent No. 5432264  
APPLICANT: GRUBB, ANDERS; LUNDWALL, AKE; ABRAHAMSON, MAGNUS;  
DALBOGE, HENRIK  
TITLE OF INVENTION: RECOMBINANT 3-DES-OH-CYSTATIN C PRODUCED  
BY EXPRESSION IN A PROCARYOTIC HOST CELL  
NUMBER OF SEQUENCES: 8

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/929,290  
FILING DATE: 13-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 440,221  
FILING DATE: 21-NOV-1989  
APPLICATION NUMBER: 297,198  
FILING DATE: 20-MAY-1988  
SEQ ID NO: 6  
LENGTH: 146  
5432264-6

Query Match  
Best Local Similarity 100.0%; Score 6; DB 6; Length 146;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIL 15  
Db 10 LLLAIL 15

RESULT 44  
US-09-085-761A-49  
Sequence 49, Application US/09085761A  
Patent No. 635178  
GENERAL INFORMATION:  
APPLICANT: Weiner, Joel H.  
APPLICANT: Turner, Raymond J.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PROTEIN  
NUMBER OF SEQUENCES: 77  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medien & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/085,761A  
FILING DATE: 28-MAY-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Peter G.  
REGISTRATION NUMBER: 32,837  
REFERENCE/DOCKET NUMBER: UALB-03356  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 161 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-09-085-761A-49

Query Match  
Best Local Similarity 100.0%; Score 6; DB 4; Length 161;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 OKPETT 99  
Db 140 OKPETT 145

RESULT 45

US-09-443-041A-6  
; Sequence 6, Application US/09443041A  
; Patent No. 6465717

GENERAL INFORMATION:

APPLICANT: Farnodu, Omolayo O.

APPLICANT: Orozco, Buddy

APPLICANT: Rafalski, Antoni

APPLICANT: Shen, Jennie

TITLE OF INVENTION: Sterol Metabolism Enzymes

FILE REFERENCE:

CURRENT APPLICATION NUMBER: US/09/443,041A

CURRENT FILING DATE: 1999-11-18

PRIOR APPLICATION NUMBER: 60/109,283

PRIOR FILING DATE: 1998-11-20

NUMBER OF SEQ ID NOS: 33

SOFTWARE: Microsoft Office 97

SEQ ID NO 6

LENGTH: 161

TYPE: PRT

ORGANISM: Glycine max

FEATURE:

NAME/KEY: UNSURE

LOCATION: (77)

OTHER INFORMATION: ANY AMINO ACID

NAME/KEY: UNSURE

LOCATION: (95)

OTHER INFORMATION: ANY AMINO ACID

NAME/KEY: UNSURE

LOCATION: (102)

OTHER INFORMATION: ANY AMINO ACID

NAME/KEY: UNSURE

LOCATION: (157)

OTHER INFORMATION: ANY AMINO ACID

NAME/KEY: UNSURE

LOCATION: (159)

OTHER INFORMATION: ANY AMINO ACID

US-09-443-041A-6

Query Match 4.4%; Score 6; DB 4; Length 161;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ALQULL 12

Db 79 ALQULL 84

RESULT 46

US-09-325-932A-46

; Sequence 46, Application US/09325932A

; Patent No. 6451604

GENERAL INFORMATION:

APPLICANT: Flinn, Barry

APPLICANT: Lasham, Annette

TITLE OF INVENTION: Compositions affecting programmed cell

TITLE OF INVENTION: death and their use in the modification of forestry plant develo

FILE REFERENCE: 1022

CURRENT APPLICATION NUMBER: US/09/325,932A

CURRENT FILING DATE: 1999-06-04

NUMBER OF SEQ ID NOS: 206

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 46

LENGTH: 199

TYPE: PRT

ORGANISM: Pinus radiata

US-09-325-932A-46

Query Match 4.4%; Score 6; DB 4; Length 199;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 LLAILL 16

|||||

Db 88 LLAILL 93

RESULT 47

US-08-792-019B-10

; Sequence 10, Application US/08792019B

; Patent No. 5741772

GENERAL INFORMATION:

APPLICANT: CHANG, MING-SHI

TITLE OF INVENTION: THE NEUTROTROPHIC FACTOR NNT-1

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSER: AMGEN INC.

STREET: 1840 DEHAVILLAND DRIVE

CITY: THOUSAND OAKS

STATE: CA

COUNTRY: USA

ZIP: 91320

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/792,019B

FILING DATE: 03-FEB-1997

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: COOK, ROBERT R.

REGISTRATION NUMBER: 31,602

REFERENCE/DOCKET NUMBER: A-442

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 204 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FEATURE:

NAME/KEY: Protein

LOCATION: 1..174

FEATURE:

NAME/KEY: Region

LOCATION: -30..0

US-08-792-019B-10

Query Match 4.4%; Score 6; DB 1; Length 204;

Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ALQULL 12

Db 14 ALQULL 19

RESULT 48

US-08-988-819-10

; Sequence 10, Application US/08988819

; Patent No. 6054294

GENERAL INFORMATION:

APPLICANT: CHANG, MING-SHI

TITLE OF INVENTION: NEUTROTROPHIC FACTOR NNT-1

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSER: AMGEN INC.

STREET: ONE AMGEN CENTER DRIVE

CITY: THOUSAND OAKS

STATE: CA

COUNTRY: USA

ZIP: 91320

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/988,819  
FILING DATE: 12-DEC-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/792,019  
FILING DATE: 03-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: COOK, ROBERT R.  
REGISTRATION NUMBER: 31,602  
REFERENCE/DOCKET NUMBER: A-442A  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 204 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..174  
FEATURE:  
NAME/KEY: Region  
LOCATION: -30..0  
US-08-988-819-10

Query Match 4.4%; Score 6; DB 3; Length 204;  
Best Local Similarity 100.0%; Pred. No. 2e+02; 0; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0;

QY 7 ALQLL 12  
Db 14 ALQLL 19

RESULT 49  
US-09-016-534-10  
Sequence 10, Application US/09016534  
Patent No. 6143874  
GENERAL INFORMATION:  
APPLICANT: CHANG, MING-SHI  
APPLICANT: ELLIOTT, GARY S.  
APPLICANT: SARMIENTO, ULIA  
APPLICANT: SENALDI, GIORGIO  
TITLE OF INVENTION: THE NEUTROPHILIC FACTOR NNT-1  
NUMBER OF SEQUENCES: 16  
CLASSIFICATION:  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: AMGEN INC.  
STREET: ONE AMGEN CENTER  
CITY: THOUSAND OAKS  
STATE: CA  
COUNTRY: USA  
ZIP: 91320  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/016,534  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/792,019  
FILING DATE: 03-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: COOK, ROBERT R.  
REGISTRATION NUMBER: 31,602  
REFERENCE/DOCKET NUMBER: A-442B  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:

LENGTH: 204 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..174  
FEATURE:  
NAME/KEY: Region  
LOCATION: -30..0  
US-09-016-534-10

Query Match 4.4%; Score 6; DB 3; Length 204;  
Best Local Similarity 100.0%; Pred. No. 2e+02; 0; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0;

QY 7 ALQLL 12  
Db 14 ALQLL 19

RESULT 50  
US-08-097-869-5  
Sequence 5, Application US/08097869  
Patent No. 6204364  
GENERAL INFORMATION:  
APPLICANT: Todaro, George J.  
APPLICANT: Rose, Timothy M.  
TITLE OF INVENTION: HYBRID CYTOKINES  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Morrison & Foerster  
STREET: 545 Middlefield Road, Suite 200  
CITY: Menlo Park  
STATE: California  
COUNTRY: USA  
ZIP: 94025  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/097,869  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/753,178  
FILING DATE: 30-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Murashige, Kate H.  
REGISTRATION NUMBER: 29,959  
REFERENCE/DOCKET NUMBER: 24455-20001.00  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-813-5600  
TELEFAX: 415-327-2951  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 204 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-097-869-5

Query Match 4.4%; Score 6; DB 3; Length 204;  
Best Local Similarity 100.0%; Pred. No. 2e+02; 0; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0;

QY 7 ALQLL 12  
Db 14 ALQLL 19

Wed Jan 21 12:38:35 2004

us-09-941-314-2.01go.ra1

Page 20

Search completed: January 21, 2004, 12:28:29  
Job time : 22 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 12:04:49 ; Search time 24.7484 Seconds  
(without alignments)  
532.362 Million cell updates/sec

Title: US-09-941-314-2  
Perfect score: 736  
Sequence: 1 MAEPWQMLQILLAILTLMA.....VFAVPPEQYKILNKSCSSD 137

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues  
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	214	29.1	139	2	A45361
2	179.5	24.4	140	2	A36163
3	174	23.6	141	2	B29632
4	173.5	23.6	139	1	UDCH
5	171.5	23.3	127	2	S07085
6	171	23.2	146	1	UDHU
7	168	22.8	120	2	S10587
8	161	21.9	141	1	UDHUP1
9	153.5	20.9	112	1	UDBO
10	153.5	20.9	141	1	UDHUP2
11	147.5	20.0	142	2	A47142
12	134	18.2	111	1	UC2040
13	134	18.2	132	2	UC4918
14	129.5	17.6	141	2	JQ1470
15	127	17.3	111	2	A28793
16	102.5	13.9	133	2	JC4536
17	98.5	13.4	433	2	KGBOL2
18	98.5	13.4	619	1	KGBOL2
19	98	13.3	436	2	TJ1871
20	96.5	13.1	436	1	KGBOL1
21	96.5	13.1	621	2	KGBOL1
22	95.5	13.0	162	2	A43428
23	93.5	12.7	427	1	KGHU1
24	93.5	12.7	644	1	KGHU1
25	87	11.8	433	2	A28051
26	87	11.8	639	2	A25486
27	85.5	11.6	430	2	A23897
28	85.5	11.6	582	2	S42613
29	81	11.0	430	2	B28055

30	79	10.7	430	1	KGR11	T-kininogen I prec
31	77.5	10.5	504	2	S54744	cellulase (EC 3.2.
32	77	10.5	430	2	D64151	hypothetical prote
33	75.5	10.3	400	2	A46297	beta-1,6-N-acetyl9
34	75.5	10.3	582	2	S42614	membrane protein P
35	75	10.2	423	1	KGR1	major acute phase
36	73.5	10.0	232	2	G95044	conserved hypother
37	73.5	10.0	438	2	A47702	glucan 1,3-beta-gl
38	73.5	10.0	438	2	T52149	beta-glucanase (im
39	73.5	10.0	519	2	A89903	conserved hypother
40	73	9.9	861	2	S12499	CH1 protein - yea
41	72.5	9.9	366	2	S14570	tubulin beta chain
42	72.5	9.9	525	1	KGHUCH	histidine-rich gly
43	72	9.8	142	2	T33301	hypothetical prote
44	72	9.8	245	2	A59258	tetraspan TSPAN-6
45	71.5	9.7	132	2	D88508	protein H1A12.5 (

## ALIGNMENTS

RESULT 1  
A45361  
cystatin-related epididymal specific protein - mouse (fragment)  
C:Species: Mus musculus (house mouse)  
C:Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 05-Nov-1999  
C:Accession: A45361  
R:Corwall, G.A.; Orgebin-Christ, M.C.; Hamm, S.R.  
Mol. Endocrinol. 6, 1653-1664, 1992  
A:Title: The CREB gene: a unique testis-regulated gene related to the cystatin family is  
A:Reference number: A45361; MUID:93078799; PMID:1280328  
A:Accession: A45361  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-139 <COR>  
A:Cross-references: GB:549926; NID:9260492; PIDN:AAC35390.1; PID:9260493  
A>Note: sequence extracted from NCBI backbone (NCBI:P:118813)  
C:Superfamily: cystatin; cystatin homology  
F:28-139/Domain: cystatin homology <CYS>

Query Match 29.1%; Score 214; DB 2; Length 139;  
Best Local Similarity 37.1%; Pred. No. 1.6e-13;  
Matches 43; Conservative 29; Mismatches 30; Indels 14; Gaps 4;

Qy 32 SYHEVMAVENY-----AKDSLOWITDOYKESDDKXHFRIFFVLKVGQVTDHLR 81  
Db 22 SKNEVKA-QNYFGSINISNANVQCWPAKMKBYKESBDKVFVLDKILHAKLQIDTDRMS 80  
Qy 82 YHLNVEKQMTTCQKP--ETNCTPQER-ELHKOVNCFPSVPAVWPEQYKILNKSC 134  
Db 81 YQIDVQISRSNCKPLNNTENCIPQKKRBLKQWCSFLVGLAPWNGFILLSKRC 136

## RESULT 2

A36163  
cystatin C precursor - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 14-Dec-1990 #sequence\_revision 14-Dec-1990 #text\_change 16-Jul-1999  
C:Accession: A36163  
R:Solom, M.; Rawson, C.; Lindburg, K.; Barnes, D.  
Biochem. Biophys. Res. Commun. 172, 945-951, 1990  
A:Title: Transforming growth factor beta regulates cystatin C in serum-free mouse embryo  
A:Reference number: A36163; MUID:91054522; PMID:2241983  
A:Accession: A36163  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-140 <SOL>  
A:Cross-references: EMBL:MS9470; NID:9192911; PIDN:AAA63298.1; PID:9192912  
C:Superfamily: cystatin; cystatin homology  
F:29-140/Domain: cystatin homology <CYS>  
F:93-103,117-137/Disulfide bonds: #statue predicted

Query Match 24.4%; Score 179.5; DB 2; Length 140;





A/Reference number: A32732; MUID:84128015; PMID:6365094  
 A/Accession: A32732  
 A/Molecule type: protein  
 A/Residues: 27-76 <BR2>  
 R/Olafsson, I.; Gudmundsson, G.; Abrahamson, M.; Jensen, O.; Grubb, A.  
 Scand. J. Clin. Lab. Invest. 50, 85-93, 1990  
 A/Title: The amino terminal portion of cerebrospinal fluid cystatin C in hereditary cyst  
 A/Reference number: A60552; MUID:90193615; PMID:2315647  
 A/Accession: A60552  
 A/Molecule type: protein  
 A/Residues: 27-49, 'XX', 52-64 <OLA>  
 A/Note: This protein, purified from cerebrospinal fluid of patients with the autosomal d  
 e defective gene is not present in CSF but is found instead in amyloid deposits  
 R/Popovic, T.; Birin, J.; Ritonja, A.; Turk, V.  
 Biol. Chem. Hoppe-Seyler 371, 575-580, 1990  
 A/Title: Different forms of human cystatin C.  
 A/Reference number: S10607; MUID:91025625; PMID:2222856  
 A/Accession: S10607  
 A/Molecule type: protein  
 A/Residues: 27-53 <POB>  
 A/Experimental source: urine, kidney disease  
 A/Note: truncated forms with amino ends at positions 35 and 36 of the precursor were als  
 R/Grubb, A.; Loiberg, H.; Barrett, A.J.  
 FEBS Lett. 170, 370-374, 1984  
 A/Title: The disulphide bridges of human cystatin C (gamma-trace) and chicken cystatin.  
 A/Reference number: S01462  
 A/Contents: annotation; disulfide bonds  
 R/Berti, P.J.; Storer, A.C.  
 Biochem. J. 302, 411-416, 1994  
 A/Title: Local pH-dependent conformational changes leading to proteolytic susceptibility  
 A/Reference number: S55305; MUID:9437969; PMID:8092991  
 A/Accession: S55305  
 A/Status: preliminary  
 A/Molecule type: protein  
 A/Residues: 27-49;106-146 <BR>  
 C/Comment: This protein is found in the post-gamma-globulin fraction of cerebrospinal fl  
 f patients with certain autoimmune diseases.  
 C/Comment: This protein is an inhibitor of cysteine proteinases and may serve an importa  
 C/Comment: A mutant cystatin C, with 94-Gln, is deposited in hereditary cerebral hemorr  
 C/Genetics:  
 A/Gene: GDB:CST3  
 A/Cross-references: GDB:119917; OMIM:105150  
 A/Map position: 20p11.2-20p11.2  
 A/Intons: 81/3; 119/3  
 C/Superfamily: cystatin; cystatin homology  
 C/Keywords: amyloid; cysteine proteinase inhibitor; extracellular protein; hydroxyprolin  
 F/1-26/Domain: signal sequence #status predicted <SIG>  
 F/27-146/Product: cystatin C #status experimental <MAT>  
 F/35-146/Domain: cystatin homology <CYS>  
 F/81-85/Region: inhibitory #status predicted  
 F/28/Modified site: hydroxyproline (Pro) (partial) #status experimental  
 F/93-109,123-143/Disulfide bonds: #status experimental

Query Match 23.2%; Score 171; DB 1; Length 146;  
 Best Local Similarity 31.5%; Pred. No. 2.4e-09;  
 Matches 46; Conservative 26; Mismatches 64; Indels 10; Gaps 4;

QY 1 MAEWMQLQILALILTLMLPYARK-----KTFISVHEWMAVENYAKDSLQMTDQY 54  
 DB 1 MAGLRAPLILALIALAVALSPAGSSPGKPRPLVCGPMDASVEEGBRDLDFAVGEY 60

QY 55 NKESDDKXHFRIFRVLKQROVTDHLEHNLVNEQMTCCK--PETNCG-VPOREHLJKO 111  
 DB 61 NKASNDVHSHALVDVPRKQIVAGVNFVDELVEGRITCTKTOPNLDCPFDHPHLKRK 120

QY 112 VNCFFSVFAVWPFQYKILNKSCSD 137  
 DB 121 AFCSFOIYAVDW-OGTMTLSKSTCOD 145

RESULT 7  
 S10587  
 cystatin C - rat

C/Species: Rattus sp. (rat)  
 C/Date: 21-Nov-1993 #sequence\_revision 03-Nov-1995 #text\_change 16-Jul-1999  
 C/Accession: S10587  
 R/Bernard, F.; Bernard, A.; Faucher, D.; Capony, J.P.; Derancourt, J.; Brillard, M.; Gauch  
 Biol. Chem. Hoppe-Seyler 371 (Suppl.), 161-166, 1990  
 A/Title: Rat cystatin C: the complete amino acid sequence reveals a site for N-glycosylat  
 A/Reference number: S10587; MUID:90380276; PMID:2400577  
 A/Accession: S10587  
 A/Status: preliminary  
 A/Molecule type: protein  
 A/Residues: 1-120 <ESN>  
 A/Note: 43-Asn was also found  
 C/Superfamily: cystatin; cystatin homology  
 F/9-120/Domain: cystatin homology <CYS>

Query Match 22.8%; Score 168; DB 2; Length 120;  
 Best Local Similarity 29.5%; Pred. No. 3.7e-09;  
 Matches 36; Conservative 27; Mismatches 45; Indels 14; Gaps 3;

QY 18 LMLPYARKKTRFISVHEWMAVENYAKDSLQMTDQYNKESDDKXHFRIFRVLKQROVT 77  
 DB 9 LLAGPQADSS-----EEGVGRALDFAVSEYKNSNDVYHRAIQVAPARKQLV 57

QY 78 DHEHYHNLVNEQMTCCKPET--TNC-VPOREHLKQVNCFFSVFAVWPFQYKILNKSC 134  
 DB 58 AGIVYIVDVEMGRITCTKSQTNLTNCPFHQPHLMRALCSFOIYSPWKGTHLTLSKSC 117

QY 135 SS 136  
 DB 118 KN 119

RESULT 8  
 UDDHPI  
 N/Alternate names: cystatin SA-III; salivary acidic protein-1  
 C/Species: Homo sapiens (man)  
 C/Date: 25-Feb-1995 #sequence\_revision 08-Feb-1996 #text\_change 16-Jul-1999  
 C/Accession: S17667; S16500; A01272; A29603; S19280; A56608  
 R/Bobek, L.A.; Aguilere, A.; Levine, M.J.  
 Biochem. J. 278, 627-635, 1991  
 A/Title: Human salivary cystatin S. Cloning, sequence analysis, hybridization in situ and  
 A/Reference number: S17667; MUID:91378918; PMID:1898352  
 A/Accession: S17667  
 A/Molecule type: mRNA  
 A/Residues: 1-141 <BOB>  
 A/Cross-references: EMBL:X54667; NID:G30365; PIDN:CA38478.1; PID:G30366  
 R/Lamkin, M.S.; Jensen, J.L.; Setayesh, M.R.; Troxler, R.F.; Oppenheim, F.G.  
 Arch. Biochem. Biophys. 288, 664-670, 1991  
 A/Title: Salivary cystatin SA-III: a potential precursor of the acquired enamel pellicle.  
 A/Reference number: S16500; MUID:91378515; PMID:1898055  
 A/Accession: S16500  
 A/Status: preliminary  
 A/Molecule type: protein  
 A/Residues: 21-134, 'D', 136-141 <IHU>  
 R/Isemura, S.; Saitoh, E.; Sanada, K.  
 J. Biochem. 96, 489-498, 1984  
 A/Title: Isolation and amino acid sequence of SP-1, an acidic protein of human whole sal  
 A/Reference number: A91985; MUID:85054716; PMID:6501254  
 A/Accession: A01272  
 A/Molecule type: protein  
 A/Residues: 23-134, 'D', 136-141 <ISB>  
 R/Isemura, S.; Saitoh, E.; Ito, S.; Isemura, M.; Sanada, K.  
 J. Biochem. 96, 1311-1314, 1984  
 A/Title: Cystatin S: a cysteine proteinase inhibitor of human saliva.  
 A/Reference number: A91981; MUID:85104877; PMID:6394600  
 A/Contents: annotation; inhibitor specificity  
 R/Hawke, D.H.; Yuen, P.M.; Wilson, K.J.; Hunkapiller, M.W.  
 Biochem. Biophys. Res. Commun. 145, 1248-1253, 1987  
 A/Title: Identification of a long form of cystatin from human saliva by rapid microbore  
 A/Reference number: A29603; MUID:8270697; PMID:3496880  
 A/Accession: A29603

A:Molecule type: protein  
 A:Residues: 21-51 <HAM>  
 R:Ramaubhu, N.; Reddy, M.S.; Bergey, E.J.; Haraszthy, G.G.; Soni, S.D.; Levine, M.J.  
 Biochem. J. 280, 341-352, 1991  
 A:Title: Large-scale purification and characterization of the major phosphoproteins and  
 A:Reference number: S19279; MUID:92082469; PMID:1747107  
 A:Accession: S19280  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 21-55 <RAM>  
 R:Johnsson, M.; Richardson, C.F.; Bergey, E.J.; Levine, M.J.; Nancollas, G.H.  
 Arch. Oral Biol. 36, 631-636, 1991  
 A:Title: The effects of human salivary cystatins and statherin on hydroxyapatite crystal  
 A:Reference number: A56608; MUID:92074898; PMID:1741693  
 A:Accession: A56608  
 A:Molecule type: protein  
 A:Residues: 21-36 <JOU>  
 A:Note: sequence extracted from NCBI backbone (NCBI:P.67866)  
 A:Note: authors designate form without phosphate as cystatin S and form containing one pi  
 C:Comment: This protein strongly inhibits papain and ficin, partially inhibits stem brom  
 competitively.  
 C:Genetics:  
 A:Gene: GDB:CST4  
 A:Cross-references: GDB:136381  
 A:Map position: 20p11.2-20p11.2  
 C:Superfamily: cystatin; cystatin homology  
 C:Keywords: cysteine proteinase inhibitor; phosphoprotein; saliva  
 F:1-20/Domain: signal sequence #status predicted <SIG>  
 F:21-141/Product: cystatin S #status predicted <MAT>  
 F:30-141/Domain: cystatin homology <CYS>  
 F:76-80/Region: inhibitory #status predicted  
 F:94-104,118-138/Disulfide bonds: #status predicted

Query Match 21.9%; Score 161; DB 1; Length 141;  
 Best Local Similarity 27.5%; Pred. No. 2,1e-08;  
 Matches 39; Conservative 31; Mismatches 65; Indels 4; Gaps 3;

OY 1 MABPMQALLALITLTMALPYQARKTFL-SVEVMAVENYAKSLQWTTDOYKESD 59  
 DB 1 MABPCLTLLMLMTLAGALSSKEENRIIPGCIYDADLNDEWQRALHFAISSEYKATE 60  
 OY 60 DKHFRIFRILKQVQRTDHLHYLNTMEMQWTTQK--PETTNCVPER-ELHKQVNCFF 116  
 DB 61 DEYYRRLQVLRABEQFGVNPFDVEGRTICTKSQPNLIDTCAFHQPELQKKQCSF 120  
 OY 117 SVFAPVPEQYKILNKSC 134  
 DB 121 EIVFVPEWEDMSLVNSRC 138

RESULT 9  
 UNBO  
 Cystatin - bovine  
 N:Alternate names: thiol proteinase inhibitor  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 28-Feb-1986 #sequence\_revision 28-Feb-1986 #text\_change 06-Dec-1996  
 C:Accession: A01271  
 R:Hirado, M.; Tanasawa, S.; Sakiyama, F.; Minobe, M.; Fujii, S.  
 FEBS Lett. 186, 41-45, 1985  
 A:Title: Complete amino acid sequence of bovine colostrum low-M-r cysteine proteinase in  
 A:Reference number: A01271; MUID:85231205; PMID:3891407  
 A:Accession: A01271  
 A:Molecule type: protein  
 A:Residues: 1-112 <HIR>  
 C:Superfamily: cystatin; cystatin homology  
 C:Keywords: colostrum; cysteine proteinase inhibitor  
 F:2-112/Domain: cystatin homology <CYS>  
 F:46-52/Region: inhibitory #status predicted  
 F:66-76,90-110/Disulfide bonds: #status predicted

Query Match 20.9%; Score 153.5; DB 1; Length 112;  
 Best Local Similarity 30.6%; Pred. No. 8.5e-08;  
 Matches 30; Conservative 23; Mismatches 42; Indels 3; Gaps 2;

OY 40 ENYAKSLQWITTOYKNSPDKXFRIFRYLAKVQVTDHLHYLNTMEMQWTTQKPEPT 99  
 DB 13 EBGVEKLSFVASEFENKRSNDATQSRVRRVRRKQVSGMNTFLDVELGRTTCKSQAN 72  
 OY 100 --NC-VPQREHLKQVNCFPSVAPVPEQYKILNKSC 134  
 DB 73 LDSCPFHNPFLHREKCKSFQYVVPVPMNTINLVKFSK 110

RESULT 10  
 UNHUP2  
 Cystatin SN precursor [validated] - human  
 N:Alternate names: cystatin SA-I  
 C:Species: Homo sapiens (man)  
 C:Date: 28-May-1986 #sequence\_revision 08-Feb-1996 #text\_change 08-Dec-2000  
 C:Accession: A28110; S02489; A29632; A01273; S19279  
 R:Al-Hashimi, I.; Dickinson, D.P.; Levine, M.J.  
 J. Biol. Chem. 263, 9381-9387, 1988  
 A:Title: Purification, molecular cloning, and sequencing of salivary cystatin SA-I.  
 A:Reference number: A28110; MUID:88243825; PMID:2837486  
 A:Accession: A28110  
 A:Molecule type: mRNA  
 A:Residues: 1-141 <ALH>  
 A:Cross-references: GB:303870; NID:g337751; PIDN:AA60299.1; PID:g337752  
 R:Salton, E.; Isemura, S.; Sanada, K.; Kim, H.S.; Smithies, O.; Maeda, N.  
 Biol. Chem. Hoppe-Seyler 369, 191-197, 1988  
 A:Title: Cystatin superfamily. Evidence that family II cystatin genes are evolutionarily  
 A:Reference number: S02489; MUID:89076505; PMID:3202964  
 A:Accession: S02489  
 A:Status: not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 21-141 <SA2>  
 R:Salton, E.; Kim, H.S.; Smithies, O.; Maeda, N.  
 Gene 61, 329-338, 1987  
 A:Title: Human cysteine-proteinase inhibitors: nucleotide sequence analysis of three mem  
 A:Reference number: A91589; MUID:88185836; PMID:3446578  
 A:Accession: A29632  
 A:Molecule type: DNA  
 A:Residues: 1-86, 118-141 <SA1>  
 R:Isemura, S.; Salton, E.; Sanada, K.  
 FEBS Lett. 198, 145-149, 1986  
 A:Title: Characterization of a new cysteine proteinase inhibitor of human saliva, cystat  
 A:Reference number: A01273; MUID:86164938; PMID:3514272  
 A:Accession: A01273  
 A:Molecule type: protein  
 A:Residues: 29-141 <ISR>  
 R:Ramaubhu, N.; Reddy, M.S.; Bergey, E.J.; Haraszthy, G.G.; Soni, S.D.; Levine, M.J.  
 Biochem. J. 280, 341-352, 1991  
 A:Title: Large-scale purification and characterization of the major phosphoproteins and  
 A:Reference number: S19279; MUID:92082469; PMID:1747107  
 A:Accession: S19279  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 21-55 <RAM>  
 C:Comment: Human saliva appears to contain several cysteine proteinase inhibitors that a  
 ences. Cystatin SN, with a pi of 7.5, is a much better inhibitor of papain and dipeptid  
 C:Genetics:  
 A:Gene: GDB:CST1  
 A:Cross-references: GDB:119815; OMIM:123855  
 A:Map position: 20p11.2-20p11.2  
 C:Superfamily: cystatin; cystatin homology  
 C:Keywords: cysteine proteinase inhibitor; extracellular protein; saliva  
 F:1-20/Domain: signal sequence #status predicted <SIG>  
 F:21-141/Product: cystatin SA-I #status experimental <MAT>  
 F:30-141/Domain: cystatin homology <CYS>  
 F:76-80/Region: inhibitory #status predicted  
 F:94-104,118-138/Disulfide bonds: #status predicted

Query Match 20.9%; Score 153.5; DB 1; Length 141;  
 Best Local Similarity 27.9%; Pred. No. 1.1e-07;  
 Matches 38; Conservative 30; Mismatches 61; Indels 7; Gaps 4;

QY	6	QALDILLAILTL-MAPEYQARKKTFU---SVHEVMAVENAKDSLQWITDYAKNESDJK	61
Db	3	QHSTLTLLTLLTLLALAMSPREEDRIIIPGSIYNNDLNDEWQRLHAIASEYNKATKD	62
QY	62	YHFRIFRLATQROVTDHLEYLINVENOMTTCOK--PETNCFPOER-ELHKOVNCFPSV	118
Db	63	YTRPEPLRRLRROQTGVGVNTPFDEVVGRTICTKSGPMLDTCARHQPFLQKOLCSPEI	122
QY	119	FAVPMFEQYKILINKSC	134
Db	123	YEVEPMENRSLVKSRC	138

RESULT 11  
A47142  
cystatin D precursor - human  
C/Species: Homo sapiens (man)  
C/Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 16-Jul-1999  
C/Accession: A47142, S18212  
R/Freife, J.P., Balbin, M.; Abrahamson, M.; Velasco, G.; Dalboge, H.; Grubb, A.; Lopez-Otin, C.  
J. Biol. Chem. 268, 15737-15744, 1993  
A/Title: Human cystatin D, cDNA cloning, characterization of the *Escherichia coli* expressed  
A/Reference number: A47142; MUID:93340179; PMID:8340398  
A/Accession: A47142  
A/Status: Preliminary  
A/Molecule type: mRNA  
A/Residues: 1-142 <FR>  
A/Cross-references: GB:X10377; MID:g398710; PIDD:CAA4938.1; PID:g398711  
A/Note: single residue difference between this report and S18218 was investigated and shown to be a typo.  
R/Freife, J.P., Abrahamson, M.; Olafsson, I.; Velasco, G.; Grubb, A.; Lopez-Otin, C.  
J. Biol. Chem. 266, 20538-20543, 1991  
A/Title: Structure and expression of the gene encoding cystatin D, a novel human cysteine proteinase inhibitor  
A/Reference number: S18212; MUID:92041895; PMID:1939105  
A/Accession: S18212  
A/Status: Preliminary  
A/Molecule type: DNA  
A/Residues: 1-45, 'C', 47-142 <FR2>  
A/Cross-references: EMBL:X59664; MID:g30263; PIDD:CAA42590.1; PID:g30264  
A/Genetics:  
A/Gene: GDB:GSTS  
A/Cross-references: GDB:136380, OMIM:123858  
A/Map position: 20p11.21-20p11.21  
A/Intons: 77/3; 115/3  
C/Superfamily: cystatin  
C/Keywords: cysteine proteinase inhibitor; extracellular protein; saliva  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:30-142/Domain: cystatin homology <CTS>

	Query Match	20.0%	Score 147.5;	DB 2;	length 142;	
	Best Local Similarity	28.1%	Pred. No. 4,2e-07;			
	Matches	39;	Conservative	27;	Mismatches 68;	Indels 5; Gaps 4
QY	1	MAEPWQALQLLAILLTLMALPYQARKFTPL-SVHEVAVENAYKADSLQMTDQYNKE-S	58			
		: : : : : :				
Dd	1	MMMPHTPLLLTLTALMVAVASASQSTLAGIHATDLNDKS YGRALDFALSEYKNVILN	60			
		: : : : : : : : : :				
QY	59	DDKHFRIFRYLVAKORQVTDHLEHLINBEMQTTOCK--PETTNC-VPOERELHKVANCE	115			
		: : : : : : : : : :				
Dd	61	KDEYYSRRLQVMAAQQOIVGVGNVNYFNFKRTCTCKSPINDNCPFNDQPKLKEEFCS	120			
		: : : : : : : : : :				
QY	116	FSVFAPVPMEQYKXILNKSC	134			
		: : : : : : : : :				
Dd	121	FQINEVPMEEDKISILNYKC	139			

```

RESULT 12
JC2040
cystatin - chum salmon
N:Alternate names: cysteine proteinase inhibitor
C:Species: Oncorhynchus keta (chum salmon)
C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 16-Jul-1995
Accession: JC2040

```

R:Koide, Y.; Noso, T.  
BioSci. Biotechnol. Biochem. 58, 164-169, 1994  
A:Title: The complete amino acid sequence of pituitary cystatin from chum salmon.  
A:Reference number: JC2040; MOID:94162738, PMID:7764512  
A:Accession: JC2040  
A:Molecule type: protein  
A:Residues: 1-111 <KOI>  
A:Comment: The intracellular role of this protein is the inhibition of intralysosomal proteolysis.  
C:Superfamily: cystatrin; cystatin homology  
C:Keywords: cysteine proteinase inhibitor  
F:2-111/Domain: cystatin homology <CYS>  
F:48-52/Region: inhibitory  
F:89-109/Disulfide bonds: #status experimental

	Query Match	18.2%	Score 134;	DB 1;	length 111;
	Best Local Similarity	27.4%	Pred. No. 6.4e-06;		
	Matches	26;	Conservative	27;	Mismatches 40; Indels 2; Gaps 1;
Oy	44	KDSLQWITDQYKNSDDKYHFRIPLVAVKVQRQTVDHLEHYLVNEMOWTTCOK--PFTNC	101		
		:::::>:::::>:::::>:::::>:::::>:::::>:::::>:::::>:::::>:::::>:::::>:::::>:::::>:::::>:::::>:::::>:::::			
Dd	17	RDAIQFAVEHNKKNTNDMFPAQVAKVNNAQGVSGMKYIFTVMGRTPRCKGVEIKCS	76		
Oy	102	VPOERELHKQVNCFFSFVAAPVFPEQYLINLKSCS	136		
		:::::>:::::>:::::>:::::>:::::>:::::>:::::>:::::>:::::>:::::>:::::>:::::>:::::>:::::>:::::>:::::			
Dd	77	VHKDPMAVPFYKCTFEVMSIIPMSGIGMKNVQCQS	111		

RESULT 13  
 JC4918  
 Cystatin precursor - chum salmon  
 C1Species: *Oncorhynchus keta* (chum salmon)  
 C1Date: 26-Sep-1996 #sequence\_\_revision 01-Nov-1996 #text\_change 21-Jan-2000  
 C1Accession: JC4918  
 R1Yamashita, M.; Konagaya, S.  
 J1Biochem. 120, 483-487, 1996  
 A1Title: Molecular cloning and gene expression of chum salmon cystatin.  
 A1Reference number: JC4918; NUID:97058289; PMID:8902609  
 A1Accession: JC4918  
 A1Status: preliminary  
 A1Molecule type: mRNA  
 A1Residues: 1-132 <YAM>  
 A1Cross-references: DDBJ:D86628  
 A1Experimental source: liver  
 C1Comment: This protein is a homolog of mammalian cystatin C, and acts as an extracellular  
 C1Superfamily: cystatin; cystatin homology  
 C1Keywords: liver  
 F11-21/Domain: signal sequence #status predicted <SIG>  
 F123-132/Domain: cystatin homology <CYS>

Query Match 18.2%; Score 134; DB 2; Length 132;  
 Best Local Similarity 23.8%; Pred. No. 7.7e-06;  
 Matches 34; Conservative 32; Mismatches 59; Indels 18; Gaps 3;

```

QY      1 MAEPQALQLLALILT-----LMAEPYQARKKTFLSVHEVMAVENTAKSLQMTIQYN 55
DB      1 MIMEMKIVPLLVAFTVANNAGLIGPDDANN-----DQGRTRDQFAVVEHN 49

QY      56 KESDDKTHFRIFRYLTKQROYTDLEHNLAVEMQTTQCK--PETTNCVYPERELHAKVN 113
DB      50 KKTIDMFRCQAKAVNNAQKQVSSGMKYLFTVQMGRTPCRKGVKICSVHNDPQNAVYK 109

QY      114 CFSVPAVPMPEQYKLIINKSSCS 136
DB      110 CTFEVMSSRFPMMSDIQVANKQCES 132
  
```

RESULT 14  
J01470  
cytatelin S precursor - rat  
C|Species: Rattus norvegicus (Norway rat)  
C|Date: 17-Jul-1992 #sequence revision 17-Jul-1992 #text\_change 16-Jul-1995  
C|Accession: J01470; A31891; 505252; S15132  
R|Cox, J.L.; Shaw, P.A.

Gene 110, 175-180, 1992

A&gt;Title: Structure, organization and regulation of a rat cysteine proteinase inhibitor-c

A/Reference number: J01470; MUID:92165056; PMID:1537554

A/Accession: J01470

A/Molecule type: DNA

A/Residues: 1-141 &lt;COX&gt;

A/Cross-references: GB:M75281; NID:g294537; PIDN:AAA1068.1; PID:g294538

R/Shaw, P.A.; Cox, J.L.; Barka, T.; Naito, Y.

J. Biol. Chem. 263, 18133-18137, 1988

A&gt;Title: Cloning and sequencing of cDNA encoding a rat salivary cysteine proteinase inh

A/Reference number: A31891; MUID:89053983; PMID:3263967

A/Accession: A31891

A/Molecule type: mRNA

A/Residues: 151DYLYISFEHETLSC', 27-141 &lt;SHA&gt;

R/Bedl, G.S.

Arch. Biochem. Biophys. 273, 245-253, 1989

A&gt;Title: Amino acid sequence of an inducible cysteine proteinase inhibitor (cystatin) fr

A/Reference number: S05252; MUID:89334379; PMID:2757396

A/Accession: S05252

A/Molecule type: protein

A/Residues: 28-113, 'QE', 116-141 &lt;BED&gt;

A/Note: 87-Glu and 88-His were also found; disulfide bonds determined

R/Nishitani, T.; Ishibashi, K.; Abe, K.

Biochim. Biophys. Acta 1077, 346-354, 1991

A&gt;Title: Isolation of three forms of cystatin from submandibular saliva of isoproterenol

A/Reference number: S15132; MUID:91230145; PMID:1903068

A/Accession: S15132

A/Molecule type: Protein

A/Residues: 29-38;76-80 &lt;BIO&gt;

A/Note: forms RSC-1, RSC-2 and RSC-3 with differing amino-terminals were found; a form h

C/Genetics:

A/Intons: CysS

A/Intons: 76/3; 114/3

C/Superfamily: cystatin; cystatin homology

C/Keywords: cysteine proteinase inhibitor

F/1-28/Domain: signal sequence #status predicted &lt;SIG&gt;

F/28-141/Product: cystatin S, form RSC-3 #status predicted &lt;MAT3&gt;

F/29-141/Product: cystatin S, form RSC-2 #status predicted &lt;MAT2&gt;

F/30-141/Product: cystatin S, form RSC-1 #status predicted &lt;MAT1&gt;

F/32-141/Product: cystatin S, form RSC-1 #status predicted &lt;MAT1&gt;

F/94-104,118-138/Disulfide bonds: #status experimental

Query Match 17.6%; Score 129.5; DB 2; Length 141;

Best Local Similarity 31.7%; Pred. No. 2.3e-05;

Matches 32; Conservative 20; Mismatches 44; Indels 5; Gaps 3;

Qy 40 ENYAKSLQWITDQYNKESDDKXFRIFRYLAKVQROYTDHLEHLYANEMQTTQOK--E 97

Db 41 EEGSESLNVAVNEYNKNSDLYLRSVEVDVQKQVAGTKFFPDVILKTLCKTQGD 100

Qy 98 TTNCVPOERELHKQVN--CFPSVFAVPMPFQYKILNKSCS 136

Db 101 LTNC-PLNERADQOEHEFCSTVVDHIDPENTIVLSSSCHS 140

RESULT 15

A28793

Cystatin - puff adder

C/Species: Bitis arietans (puff adder)

C/Date: 15-Dec-1988 #sequence\_revision 15-Dec-1988 #text\_change 30-Sep-1993

C/Accession: A28793

R/Ritonja, A.; Evans, H.J.; Machleidt, W.; Barrett, A.J.

Biochem. J. 246, 799-802, 1987

A&gt;Title: Amino acid sequence of a cystatin from venom of the African puff adder (Bitis a

A/Reference number: A28793; MUID:88076861; PMID:3500714

A/Accession: A28793

A/Molecule type: protein

A/Residues: 1-111 &lt;RT&gt;

C/Superfamily: cystatin; cystatin homology

Query Match 17.3%; Score 127; DB 2; Length 111;

Best Local Similarity 30.3%; Pred. No. 3e-05;

Matches 27; Conservative 24; Mismatches 28; Indels 10; Gaps 3;

Qy 44 KDSLQWITDQYNKESDDKXFRIFRYLAKVQROYTDHLEHLYANEMQTTQOK----- 95

Db 16 QEAAAFVBEKYNAGSRNDYFKERFVBEAOSQVAGSVKYLMMELTKTKTVGRPKGY 75

Qy 96 PETTNC-VPOERELHKQVNCFPSVFAVPW 123

Db 76 QETQNCPLPENO-QEITRFVWSRPW 103

Search completed: January 21, 2004, 12:09:16

Job time : 26.7484 secs

THIS PAGE BLANK (user)

THIS PAGE BLANK (user)

BEST AVAILABLE COPY

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 12:04:04 ; Search time 54.8 Seconds  
(without alignments)  
645.132 Million cell updates/sec

Title: US-09-941-314-2

Perfect score: 1 MAEPWQALQLLAILTLTMA.....VPAVPEQYKILNKSCSSD 137

Sequence:

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL 23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mmc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriophage:\*
- 17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	736	100.0	138	4	Q8WXU6
2	507.5	69.0	103	4	Q8WXU5
3	404	54.9	139	11	Q8K5A3
4	192.5	26.2	140	11	Q9BFX9
5	191	26.0	141	11	Q9DAP1
6	139	18.9	112	13	Q98SR4
7	139	18.9	112	13	Q98SR3
8	123	16.7	148	11	Q9JMH4
9	122	16.6	130	11	Q8VTH8
10	121	16.4	130	11	Q9CK46
11	121	16.4	130	11	Q8VTH3
12	116.5	15.8	128	11	Q9DAN8
13	114	15.5	149	11	Q8VHC1
14	112.5	15.3	146	11	Q8K397
15	112.5	15.3	149	11	Q9DIB1
16	108	14.7	167	11	Q9QWL5

17	105	14.3	128	11	Q8VTH2
18	101	13.7	133	11	Q9D264
19	100.5	13.7	144	13	Q8JF05
20	98	13.3	498	5	Q16454
21	95.5	13.0	109	5	Q9TY65
22	92.5	12.6	161	5	Q16159
23	90.5	12.3	148	5	Q9NH95
24	85.5	11.6	582	2	Q49557
25	84.5	11.5	125	5	Q25620
26	84.5	11.5	127	5	P90698
27	84	11.4	425	3	Q12700
28	83.5	11.3	1824	5	Q9NBH3
29	82.5	11.2	157	5	Q17108
30	82	11.1	127	5	Q9U9A1
31	81.5	11.1	115	6	Q95K43
32	80.5	10.9	133	5	Q8MVB6
33	79.5	10.8	2773	5	Q8IBF4
34	79	10.7	209	5	Q8ISX0
35	79	10.7	430	11	Q63581
36	79	10.7	432	11	Q91XK5
37	79	10.7	576	17	Q8UD0D
38	78.5	10.7	159	4	Q8TD53
39	78	10.6	996	4	Q8NDM7
40	76.5	10.4	406	5	Q8IM32
41	75.5	10.3	400	4	Q8WKK6
42	75.5	10.3	400	4	Q8WKK7
43	75.5	10.3	410	10	Q9FFW2
44	75.5	10.3	582	2	Q49558
45	75.5	10.3	5415	5	Q8IB74

## ALIGNMENTS

### RESULT 1

Q8WXU6 PRELIMINARY; PRT; 138 AA.

AC Q8WXU6;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE SC13.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
CX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N. A.  
RA Hamil K.G., Liu Q., Zhang Y.-L., French P.S., Hall S.H.;  
RT "SC13: A novel epididymal specific member of the cystatin family."  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF35480; AAL7191.1; -  
DR InterPro: IPR000010; Cystatin.  
DR Pfam; PR00031; Cystatin; 1.  
DR SMART; SM00043; CT; 1.  
SQ SEQUENCE 138 AA; 16506 MW; E49440ACA3585C64 CRC64;

Query Match	Similarity	100.0%	Score 736	DB 4	Length 138
Best Local	Similarity	100.0%	Pred. No. 1.5e-69		
Matches 137	Conservative	0	Mismatches 0	Indels 0	Gaps 0
QY	1	MAEPWQALQLLAILTLTMA	LPYARKKTFPSVSRVNA	AVENYAKDSQWITDQYNKSSD	60
Db	2	MAEPWQALQLLAILTLTMA	LPYARKKTFPSVSRVNA	VENYAKDSQWITDQYNKSSD	61
QY	61	KYHRIFRVLKVQRQVTDHLE	YHNVEMQWTCXPEFTNCV	POBRILHKVNCFFSYFA	120
Db	62	KYHRIFRVLKVQRQVTDHLE	YHNVEMQWTCXPEFTNCV	POBRILHKVNCFFSYFA	121
QY	121	VPWFQYKILNKSSD	137		
Db	122	VPWFQYKILNKSSD	138		

## RESULT 2

Q8WKX5 PRELIMINARY; PRT; 103 AA.  
 AC Q8WKX5; 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE SC3delta.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hamil K.G., Liu Q., Zhang Y.-L., French F.S., Hall S.H.;  
 RT "SC13: A novel epididymal specific member of the cystatin family";  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF35481; AAL71992.1; -  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF00031; Cystatin; 1.  
 DR SMART; SM00043; Cystatin; 1.  
 SQ SEQUENCE 103 AA; 12285 MW; 05DD92C47387B022 CRC64;

Query Match 69.0%; Score 507.5; DB 4; Length 103;  
 Best Local Similarity 74.5%; Pred. No. 1e-45;  
 Matches 102; Conservative 0; Mismatches 0; Indels 35; Gaps 1;

QY 1 MAEPWQALQILLALITLTMALPYQARKKTFSLVHEWAVENYADSLQWITDQNKESDD 60  
 DB 2 MAEPWQALQILLALITLTMALPYQARKKTFSLVHEWAVENYADSLQWITDQNKESDD 61  
 QY 61 KYHFRIRFVLKVQROVTDHLEHNVEMQWTTCCQKPTTNCVPOREILHKQVNCFFSVFA 120  
 DB 62 KYHFRIRFVLKVQROVTDHLEHNVEMQWTTCCQKPTTNCVPOREILHKQVNCFFSVFA 86  
 QY 121 VPMFEQYKIINKSCSSD 137  
 DB 87 VPMFEQYKIINKSCSSD 103

## RESULT 3

Q8KSA3 PRELIMINARY; PRT; 139 AA.  
 AC Q8KSA3; 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Cystatin 11.  
 GN CS711.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCB1\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RA Hamil K.G., Hall S.H.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF501290; AAM21709.1; -  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF00031; Cystatin; 1.  
 DR SMART; SM00043; CY; 1.  
 SQ SEQUENCE 139 AA; 16686 MW; E1E36DB786B4D08C CRC64;

Query Match 54.9%; Score 404; DB 11; Length 139;  
 Best Local Similarity 50.0%; Pred. No. 1.1e-34;  
 Matches 68; Conservative 36; Mismatches 32; Indels 0; Gaps 0;

QY 1 MAEPWQALQILLALITLTMALPYQARKKTFSLVHEWAVENYADSLQWITDQNKESDD 60  
 DB 2 MAEPWQALQILLALITLTMALPYQARKKTFSLVHEWAVENYADSLQWITDQNKESDD 61  
 QY 61 KYHFRIRFVLKVQROVTDHLEHNVEMQWTTCCQKPTTNCVPOREILHKQVNCFFSVFA 120

DB 62 LYNFRIRLIRIKIEKQWNTNHEFHITVEMQRTTCKTEKNLCNVQEGELHKQIQCFVSIVV 121  
 QY 121 VPMFEQYKIINKSCSS 136  
 DB 122 IPWLEVFEMKLNCTN 137

## RESULT 4

Q9EPX9 PRELIMINARY; PRT; 140 AA.  
 AC Q9EPX9; 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Cystatin C.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c;  
 RX MEDLINE=21010502; PubMed=11144350;  
 RA Taupin P.J., Ray J., Fischer W.H., Suhr S.T., Hakansson K., Grubb A.,  
 RA Gage F.H.;  
 RT "RGF-2-Responsive neural stem cell proliferation requires CCG, a novel  
 RT autocytine/paracrine cofactor";  
 RL Neuron 28:385-397(2000).  
 DR EMBL; AF311741; AAG40283.1; -  
 DR HSSP; P01034; I966.  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF00031; Cystatin; 1.  
 DR SMART; SM00043; CY; 1.  
 DR PROSITE; PS00287; CYSTATIN; 1.  
 FT CHAIN 21 140 CYSTATIN C.  
 FT VARIANT 16 16 A -> G.  
 FT VARIANT 84 84 L -> F.  
 SQ SEQUENCE 140 AA; 15517 MW; 3A563406D58D785 CRC64;

Query Match 26.2%; Score 192.5; DB 11; Length 140;  
 Best Local Similarity 31.7%; Pred. No. 1.7e-12;  
 Matches 44; Conservative 30; Mismatches 62; Indels 3; Gaps 2;

QY 1 MAEPWQALQILLALITLTMALPYQARKKTFSLVHEWAVENYADSLQWITDQNKESDD 60  
 DB 1 MASPRLSLLELVALAVAMATPKQGRMGADBEADANEQVRALDFAVSEYNGSND 60  
 QY 61 KYHFRIRFVLKVQROVTDHLEHNVEMQWTTCCQKPTTNCVPOREILHKQVNCFFS 117  
 DB 61 AYHSRAQVVRARQQLVAGVNYFLDVMGRTTCTKSGTNLTDCPRHQPHLMKALCSFQ 120  
 QY 118 VPMFEQYKIINKSCSS 136  
 DB 121 IYSVPMKGTSLTNFSCKN 139

## RESULT 5

Q9DAP1 PRELIMINARY; PRT; 141 AA.  
 AC Q9DAP1; 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE 1700006C19RIK protein.  
 GN 1700006C19RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;

MedLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Teshi Y.,  
 RA Atakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 RA Atakawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kodori K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Pleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,  
 RA Knehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schirral L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carinci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kaniya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mommaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilting L.,  
 RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Konteuk S.,  
 RA Hayashizaki Y.,  
 RA "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 DR EMBL; AK005665; BAB24175.1; -.  
 DR HSSP; P01038; ICBW.  
 DR MGD; MGI:191654; 1700006C19Rik.  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF00031; cystatin; 1.  
 DR SMART; SM00043; CY; 1.  
 SQ SEQUENCE 141 AA; 16811 MW; C20FA0DB8B1AC378C CRC64;

Query Match 26.0%; Score 191; DB 11; Length 141;  
 Best Local Similarity 31.9%; Pred. No. 2.5e-12;  
 Matches 43; Conservative 28; Mismatches 42; Indels 22; Gaps 4;

Qy 3 EPMQALQTLALITLTLALPYQARKTFLSVHEVMAVENYAKDSLQWITDQYNKSDKY 62  
 Db 23 EAMGSPQIV-----RPEDIPKSV-----YQHLNLYAMKEYNASNDLY 63  
 Qy 63 HFRIFRLVLYQVOTDHLRYHLNLYEMQWTTQCK--PETTNCV--DQREHLKQVNCFSVP 119  
 Db 64 NFRVVDILKSOEQITDLSLEYLYEVNARTCKKVAQDNENCLPQDDPMKMKVVCIFIVS 123  
 Qy 120 AVPMFEQYKILKNSC 134  
 Db 124 SKPMKFLKMLKKQC 138  
 RESULT 6  
 Q98SR4 PRELIMINARY; PRT; 112 AA.  
 AC Q98SR4;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Cystatin (Fragment).  
 OS Acipenser sinensis (Chinese sturgeon).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Chondrostei; Acipenseriformes; Acipenseridae;  
 OC Acipenser.  
 OC NCBI\_TaxID=61970;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA Bai J., Lao H., Ye X., Li Y., Lou J.,  
 RT "Molecular cloning and sequence analysis of cystatin cDNA from two  
 RT species of sturgeons.";  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF34610; AAK16731.1; -.  
 DR HSSP; P01038; 1A90.  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF00031; cystatin; 1.  
 DR SMART; SM00043; CY; 1.  
 DR PROSITE; PS00287; CYSTATIN; 1.  
 FT NON\_TER 1 1

SQ SEQUENCE 112 AA; 12231 MW; 48ECBFBED8A08C00 CRC64;  
 Query Match 18.9%; Score 139; DB 13; Length 112;  
 Best Local Similarity 32.0%; Pred. No. 5.6e-07;  
 Matches 33; Conservative 24; Mismatches 36; Indels 10; Gaps 3;

Qy 40 ENYAKSLQWITDQYNKSDKYHFRIFRLVLYQVOTDHLRYHLNLYEMQWTTQCK--P 96  
 Db 13 EBGVDALKPAAVAFNPNASNDMTIHRVSKVQKQVQVAGIKITVYQMGRTSCRGGA 72  
 Qy 97 ETTNC-----VPOREHLKQVNCFSVPVPMFEQYKILKNSCS 135  
 Db 73 KIELCAFDV-----ELAKISTCTFEVVSRLMIFETLVKNTCT 112

## RESULT 7

Q98SR3 PRELIMINARY; PRT; 112 AA.  
 AC Q98SR3;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Cystatin (Fragment).  
 OS Acipenser schrenckii (Amur sturgeon).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Chondrostei; Acipenseriformes; Acipenseridae;  
 OC Acipenser.  
 OC NCBI\_TaxID=111304;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA Bai J., Lao H., Ye X., Li Y., Lou J.,  
 RT "Molecular cloning and sequence analysis of cystatin cDNA from two  
 RT species of sturgeons.";  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF34611; AAK16732.1; -.  
 DR HSSP; P01038; 1A90.  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF00031; cystatin; 1.  
 DR PROSITE; PS00287; CYSTATIN; 1.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 112 AA; 12231 MW; 48ECBFBED8A08C00 CRC64;  
 Query Match 18.9%; Score 139; DB 13; Length 112;  
 Best Local Similarity 32.0%; Pred. No. 5.6e-07;  
 Matches 33; Conservative 24; Mismatches 36; Indels 10; Gaps 3;  
 Qy 40 ENYAKSLQWITDQYNKSDKYHFRIFRLVLYQVOTDHLRYHLNLYEMQWTTQCK--P 96  
 Db 13 EBGVDALKPAAVAFNPNASNDMTIHRVSKVQKQVQVAGIKITVYQMGRTSCRGGA 72  
 Qy 97 ETTNC-----VPOREHLKQVNCFSVPVPMFEQYKILKNSCS 135  
 Db 73 KIELCAFDV-----ELAKISTCTFEVVSRLMIFETLVKNTCT 112  
 RESULT 8  
 Q9JW84 PRELIMINARY; PRT; 148 AA.  
 AC Q9JW84;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE DD72 protein.  
 GN Gsr10 OR DD72.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ikegawa S., Nakamura Y.,  
 RT "DD72, a novel mouse gene implicated in the early stage of ectopic

```

RT oesification."
RL EMBL; AB036743; BAA95411.1; --
DR HSSP; P01034; I936.
DR MGD; MGI:193004; Cact10.
DR InterPro; IPR000010; Cystatin.
DR InterPro; IPR001713; StefinA.
DR Pfam; PF00031; cystatin.1.
DR PRINTS; PR00295; STEFINA.
DR SMART; SM00043; CY.1.
DR PROSITE; PS00287; CYSTATIN.1.
SQ SEQUENCE 148 AA; 16451 MW; 637534CBFCA5A179 CRC64;

Query March 16.7%; Score 123; DB 11; Length 148;
Best Local Similarity 23.2%; Pred. No. 3.7e-05;
Matches 33; Conservative 30; Mismatches 71; Indels 8; Gaps 3;

QY 1 MAEPWQALQLLAILITLALP-----YQARKTFLSYHEVAVENTAKDSLQWITDOYN 55
DB 4 LLSFPMPLAVALTLTLAVIPEASTNAEAKQVVLGVEPADPKQKVEQVKKAVRTYN 63
QY 56 KESDDKHFRIFRVLKVRQYTDHLEHNLNEMQTTQKERT--TNCVPOER-ELHKQV 112
DB 64 DMDLDYLSKPIRLMSASQVAVAGKRYVYLKIELGRTTCTESMLVDCPFNEOPDQOKRV 123
QY 113 NCFPSVFAVPWFEOYKILNKSC 134
DB 124 ICNFOINVAPWMLNKMSTNENC 145

RESULT 9
08VIT8 PRELIMINARY; PRT; 130 AA.
AC 08VIT8;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Cystatin SC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; Tissue=Testis;
RA Li Y., Friel P.J., Griswold M.D.;
RT "Molecular cloning and characterization of cystatin SC and cystatin
RT TE-1, new members of the cystatin family."
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF442205; AAL35350.1; --
DR InterPro; IPR000010; Cystatin.
DR Pfam; PF00031; cystatin.1.
DR SMART; SM00043; CY.1.
SQ SEQUENCE 130 AA; 14981 MW; 7A752359860989C9 CRC64;

Query March 16.6%; Score 122; DB 11; Length 130;
Best Local Similarity 27.5%; Pred. No. 4e-05;
Matches 30; Conservative 28; Mismatches 45; Indels 6; Gaps 3;

QY 28 KTFLSVHEWAVENYAKDSLQWITDOYNKESDDKHFRIFRVLKVRQYTDHLEHNLNVE 87
DB 23 KEFLDVTYKDL---DYFAVSEFAVAQFNDNNSSENYTRLLLEVGAAQKK-TWTMTFLMDLE 78
QY 88 MQWTTCKPEPT--NCVPOERELHKQVNCPSVFAVPWFEOYKILNKSC 134
DB 79 MGRITCKKHENHNCPLDGSSEKRVHCVQVDARPFHFTLTSTC 127

RESULT 10
09CX46 PRELIMINARY; PRT; 130 AA.
AC 09CX46;
DT 01-JUN-2001 (TREMBlrel. 17, Created)

```

```

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE 8030411P24RIK protein.
GN 8030411P24RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; Tissue=Embryonic testis;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arkawa T., Hara A., Fukunishi Y., Kono H., Aichi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kaota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiraldi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Mashio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barab G.,
RA Blake J., Botfield D., Bolunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guetlich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyshewski A., Yoshida K., Hasegawa Y., Kawai H., Kohseki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK020193; BAB32024.1; --
DR HSSP; P01034; I936.
DR MGD; MGI:1925859; 8030411P24RIK.
DR InterPro; IPR000010; Cystatin.
DR Pfam; PF00031; cystatin.1.
DR SMART; SM00043; CY.1.
SQ SEQUENCE 130 AA; 14947 MW; DD2P930B64E854 CRC64;

Query March 16.4%; Score 121; DB 11; Length 130;
Best Local Similarity 27.5%; Pred. No. 5.1e-05;
Matches 30; Conservative 28; Mismatches 45; Indels 6; Gaps 3;

QY 28 KTFLSVHEWAVENYAKDSLQWITDOYNKESDDKHFRIFRVLKVRQYTDHLEHNLNVE 87
DB 23 KEFLDVTYKDL---DYFAVSEFAVAQFNDNNSSENYTRLLLEVGAAQKK-TWTMTFLMDLE 78
QY 88 MQWTTCKPEPT--NCVPOERELHKQVNCPSVFAVPWFEOYKILNKSC 134
DB 79 MGRITCKKHENHNCPLDGSSEKRVHCVQVDARPFHFTLTSTC 127

RESULT 11
08VIT3 PRELIMINARY; PRT; 130 AA.
AC 08VIT3;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Cystatin SC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57; Tissue=Testis;
RA Li Y., Friel P.J., Griswold M.D.;
RT "Molecular cloning and characterization of cystatin SC and cystatin
RT TE-1, new members of the cystatin family."
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF440735; AAL30841.1; --

```

```

DR   InterPro: IPR000010; Cystatin.
DR   Pfam: PF00031; cystatin; 1.
DR   SMART: SM00043; Cy; 1.
SQ   SEQUENCE 130 AA; 15076 MW; DD34930B64AE58F CRC64;

Query Match
Best Local Similarity 27.5%; Pred. No. 5, 1e-05;
Matches 30; Conservative 28; Mismatches 45; Indels 6; Gaps 3;

QY 28 KTFSLVHVAENVAYAKSLQWITDQYNKESDDKTHFRIVLVKVRQVTDHLEHYLWVE 87
DB 23 KEFLDVLTKDL---DYFVAJSEFAVAQFNDDNNPEENTYKLLLEVGRAQKK-TWTMTFLMDLE 78

QY 88 MOWTTCQKPEPTT--NCVQPERELHKQVNCFFSVAVPWFQYKILNKSC 134
DB 79 MGRITCKRKHDEINHCPLQGSREKVKHCVFQVADARWPFHFTLTSTC 127

RESULT 12
ID Q9DAN8 PRELIMINARY; PRT; 128 AA.
AC Q9DAN8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE 1700006F03RIK protein (Cystatin TE-1).
GN 1700006F03RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Atakawa T., Hara A., Fukunishi Y., Komoto H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamamata I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Pleichmann T., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schirra L.M., Scandoli P., Suzuki R., Tomita M., Wagner D., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Baroni G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guenrich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker G., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohzuki S.,
RA Hayaishizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).

[2]
SEQUENCE FROM N.A.
RP STRAIN=C57; TISSUE=Testis;
RC Li Y., Friel P.J., Griswold M.D.;
RT "Molecular cloning and characterization of cystatin SC and cystatin
RT TE-1, new members of the cystatin family.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AK005670; BAB24179.1; -.
DR EMBL: AF440737; AL30843.1; -.
DR MGD: MGI:1916612; 1700006F03RIK.
DR InterPro: IPR000010; Cystatin.
DR Pfam: PF00031; cystatin; 1.
DR SMART: SM00043; Cy; 1.
SQ SEQUENCE 128 AA; 15036 MW; 40AFD00103B55BD6 CRC64;

Query Match
Best Local Similarity 25.2%; Pred. No. 0.00015;
Matches 34; Conservative 30; Mismatches 54; Indels 17; Gaps 5;

```

```

QY 5 WQALQILLAILLTLMALPYQARKKTFSLV---HEWAVENVAYAKSLQWITDQYNKESDDK 61
DB 3 WKSIVSLVALIVGLIDHCSFK-----PLEIDKNEEPRV-----SVEHVVFHRENODDD 51

QY 62 YHRIFRIVLVKVRQVTDHLEHYLWVEWMTTCQK--PETTCVQPERELHKQVNCFFSVF 119
DB 52 FAKFKPLRVRSRLQKXT-LKTVLDLGMGRLCGKQYBEDINDCPLQSGPGRKRVCTYIVE 110

QY 120 AVPFQYKILNKSC 134
DB 111 TEAMVTKFTLINSTC 125

RESULT 13
ID Q8VHC1 PRELIMINARY; PRT; 149 AA.
AC Q8VHC1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cystatin N.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;

SEQUENCE FROM N.A.
RP Hong J., Roemer M.R.;
RT "Characterization of a novel cysteine proteinase inhibitor, cystatin
RT N.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY046415; AL02328.1; -.
DR InterPro: IPR000010; Cystatin.
DR Pfam: PF00031; cystatin; 1.
DR SMART: SM00043; Cy; 1.
SQ SEQUENCE 149 AA; 16792 MW; 190DD8A89BC133C0 CRC64;

Query Match
Best Local Similarity 15.5%; Score 114; DB 11; Length 149;
Matches 38; Conservative 28; Mismatches 61; Indels 26; Gaps 6;

QY 1 MARRPQALQILLAILLTLMALPYQARKKTFSLVHVAENVAYAKD-----SLQWITDQ 53
DB 1 MERRHPPLAAGLGLAFCLTTLTSPDPAAR--LSSRRTERGKONLSPDDPRVLKAAQAAVAS 58

QY 54 YNRESDDKYFRIFRVLKVRQVTDHLEHYLWVEWMTTCQK-----ETTNC----- 101
DB 59 YNDSNSLYFRTRKYIDAKCQLVAGIKYMYVDISTECKTRVSGDHDLTTCPLAAG 118

QY 102 VQPERELHKQVNCFFSVAVPWFQYKILNKSC 134
DB 119 VQCEK-----LRGNFELBLVPMKWTQLLKHDC 146

RESULT 14
ID Q8K397 PRELIMINARY; PRT; 146 AA.
AC Q8K397;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE RKEN cDNA 110017811 gene (fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

SEQUENCE FROM N.A.
RP STRAIN=Mammary gland;
RC Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC027680; AAH27680.1; -.

```

DR InterPro: IPR000010; Cystatin.  
 DR Pfam: PF00031; Cystatin; 1.  
 DR SMART: SM00043; CY; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 146 AA; 16380 MW; 9D7BB9A6063A5C4 CRC64;  
 Query Match 15.3%; Score 112.5; DB 11; Length 146;  
 Best Local Similarity 26.1%; Pred. No. 0.00046;  
 Matches 23; Conservative 20; Mismatches 38; Indels 7; Gaps 1;  
 QY 54 YNKSDDKYHFRIFRVLVKQROVTDHLEHYNEMQWTTCKP-----ETTNCVPOER 106  
 DB 56 YNMSDSLYYFRDVKVDAKYQLVAGIKYLTLDIESTECKRVRSGEHMDLTTCPLAAG 115  
 QY 107 ELHQVNCFFSVAPVWPEQYKILNKS 134  
 DB 116 GQOEKLRNCFELLEVPMKNTTQLKHDC 143  
 RESULT 15  
 Q9DIB1 PRELIMINARY; PRT; 149 AA.  
 AC Q9DIB1;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE 1110017E1IRK protein (Cystatin M/E) (Cystatin N homolog).  
 GN 1110017E1IRK OR C3T6.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryo;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,  
 RA Saito T., Okazaki Y., Gojodori T., Bono H., Kasukawa T., Saito K.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schirrl L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Botfield D., Boujunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Guenichich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombereis P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyshaw-Boris A., Yoshida K., Haasegawa Y., Kawaji H., Kohlschki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129S6/SvEvTac; TISSUE=Spleen;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA Zeeuwen P.L.J.M., van Vlijmen-Willems I.M.J.J., Hendriks W.,  
 RA Merckx G.F., Schaikwijk J.;  
 RT "A mouse cystatin M/E-null mutation.";  
 RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 DR EMBL; AK003744; BAB22976.1; -

DR EMBL; AY093591; AAM11475.1; -  
 DR EMBL; AK078116; BAC37132.1; -  
 DR HSSP; P01038; ICEW  
 DR MED; MG1:1920970; Cst6.  
 DR InterPro: IPR000010; Cystatin.  
 DR Pfam: PF00031; Cystatin; 1.  
 SQ SEQUENCE 149 AA; 16796 MW; E713BB920E0FCC5 CRC64;  
 Query Match 15.3%; Score 112.5; DB 11; Length 149;  
 Best Local Similarity 26.1%; Pred. No. 0.00047;  
 Matches 23; Conservative 20; Mismatches 38; Indels 7; Gaps 1;  
 QY 54 YNKSDDKYHFRIFRVLVKQROVTDHLEHYNEMQWTTCKP-----ETTNCVPOER 106  
 DB 59 YNMSDSLYYFRDVKVDAKYQLVAGIKYLTLDIESTECKRVRSGEHMDLTTCPLAAG 118  
 QY 107 ELHQVNCFFSVAPVWPEQYKILNKS 134  
 DB 119 GQOEKLRNCFELLEVPMKNTTQLKHDC 146  
 Search completed: January 21, 2004, 12:08:36  
 Job time : 56.8 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: January 21, 2004, 12:08:45 ; Search time 48.6129 Seconds  
(without alignments)  
576.265 Million cell updates/sec

Title: US-09-941-314-2

Sequence: 1 MAFPMQALQILLAILTLTMA.....VFAVPEQYKIINKSCSSD 137

## Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 762491 seqs, 204481190 residues

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Published Applications AA:\*

- 1: /cgn2\_6/prodata/2/pubppaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/prodata/2/pubppaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/prodata/2/pubppaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/prodata/2/pubppaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/prodata/2/pubppaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/prodata/2/pubppaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/prodata/2/pubppaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/prodata/2/pubppaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/prodata/2/pubppaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/prodata/2/pubppaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/prodata/2/pubppaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/prodata/2/pubppaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/prodata/2/pubppaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/prodata/2/pubppaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/prodata/2/pubppaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/prodata/2/pubppaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/prodata/2/pubppaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/prodata/2/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	736	100.0	137	10	US-09-941-314-2
2	640	87.0	117	10	US-09-941-314-4
3	629	85.5	115	10	US-09-941-314-3
4	446	60.6	80	10	US-09-941-314-15
5	334	45.4	59	10	US-09-941-314-16
6	288	39.1	52	10	US-09-941-314-14
7	273	37.1	48	10	US-09-941-314-17
8	272	37.0	49	10	US-09-941-314-13
9	262	35.6	50	9	US-09-864-761-34832
10	262	35.6	50	9	US-09-864-761-48936
11	254	34.5	46	10	US-09-941-314-10
12	189	25.7	33	10	US-09-941-314-12
13	189	25.7	36	10	US-09-941-314-6
14	187	25.4	35	10	US-09-941-314-8
15	186.5	25.3	140	12	US-10-376-564-46

16	186.5	25.3	145	9	US-09-740-638-2	Sequence 2, Appl1
17	186.5	25.3	145	14	US-10-006-467-2	Sequence 2, Appl1
18	186.5	25.3	145	15	US-10-235-148-2	Sequence 2, Appl1
19	182.5	24.8	140	12	US-10-376-564-48	Sequence 48, Appl1
20	180.5	24.5	145	15	US-10-168-425-14	Sequence 14, Appl1
21	179.5	24.4	140	7	US-08-849-303-18	Sequence 18, Appl1
22	174	23.6	141	7	US-08-849-303-24	Sequence 24, Appl1
23	174	23.6	141	7	US-09-940-497-6	Sequence 6, Appl1
24	173.5	23.6	141	7	US-08-849-303-15	Sequence 15, Appl1
25	173.5	23.6	139	7	US-08-849-303-4	Sequence 4, Appl1
26	171.5	23.3	127	10	US-09-969-834-19	Sequence 19, Appl1
27	171	23.2	146	9	US-08-849-303-17	Sequence 17, Appl1
28	171	23.2	146	9	US-09-940-497-3	Sequence 3, Appl1
29	171	23.2	146	12	US-10-376-564-47	Sequence 47, Appl1
30	171	23.2	146	15	US-10-329-428-3	Sequence 3, Appl1
31	170.5	23.2	146	10	US-09-775-932-16	Sequence 16, Appl1
32	170	23.1	146	9	US-09-969-834-3	Sequence 3, Appl1
33	166.5	22.6	165	9	US-09-740-638-5	Sequence 5, Appl1
34	166.5	22.6	165	14	US-10-006-467-5	Sequence 5, Appl1
35	166.5	22.6	165	15	US-10-235-148-5	Sequence 5, Appl1
36	161	21.9	141	7	US-08-849-303-22	Sequence 22, Appl1
37	161	21.9	141	9	US-09-940-497-5	Sequence 5, Appl1
38	161	21.9	141	10	US-09-974-298-141	Sequence 141, App
39	161	21.9	141	12	US-10-241-220-77	Sequence 77, Appl1
40	160.5	21.8	121	10	US-09-775-932-8	Sequence 8, Appl1
41	160	21.7	181	12	US-10-264-048-2608	Sequence 2608, Ap
42	155	21.1	120	10	US-09-775-932-2	Sequence 2, Appl1
43	153.5	20.9	112	7	US-08-849-303-16	Sequence 16, Appl1
44	153.5	20.9	118	10	US-09-775-932-24	Sequence 24, Appl1
45	153.5	20.9	141	7	US-08-849-303-21	Sequence 21, Appl1

## ALIGNMENTS

RESULT 1  
US-09-941-314-2  
Sequence 2, Application US/09941314  
Patent No. US2002014236A1  
GENERAL INFORMATION:  
APPLICANT: ZymoGenetics, Inc.  
TITLE OF INVENTION: Mammalian Cystatin-B and Its Use to  
TITILE OF INVENTION: Inhibit Cancer Procoagulant Protein  
FILE REFERENCE: 00-81PC  
CURRENT APPLICATION NUMBER: US/09/941,314  
CURRENT FILING DATE: 2001-08-29  
PRIOR APPLICATION NUMBER: 60/230,230  
PRIOR FILING DATE: 2001-09-01  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 137  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-941-314-2

Query Match	100.0%	Score 736;	DB 10;	Length 137;
Best Local Similarity	100.0%	Pred. No. 5.8e-73;		
Matches 137;	Conservative 0;	Mismatches 0;	Gaps 0;	
Db	1	MAFPMQALQILLAILTLTMA	1	MAFPMQALQILLAILTLTMA
QY	1	MAFPMQALQILLAILTLTMA	1	MAFPMQALQILLAILTLTMA
Db	1	MAFPMQALQILLAILTLTMA	1	MAFPMQALQILLAILTLTMA
QY	1	MAFPMQALQILLAILTLTMA	1	MAFPMQALQILLAILTLTMA
Db	61	KYHFRFRVAVKVOYVDLHLYHLYNEMQWTTQCKPDTTCVQBERELHQNVCFFSVFA	120	KYHFRFRVAVKVOYVDLHLYHLYNEMQWTTQCKPDTTCVQBERELHQNVCFFSVFA
QY	61	KYHFRFRVAVKVOYVDLHLYHLYNEMQWTTQCKPDTTCVQBERELHQNVCFFSVFA	120	KYHFRFRVAVKVOYVDLHLYHLYNEMQWTTQCKPDTTCVQBERELHQNVCFFSVFA
Db	61	KYHFRFRVAVKVOYVDLHLYHLYNEMQWTTQCKPDTTCVQBERELHQNVCFFSVFA	120	KYHFRFRVAVKVOYVDLHLYHLYNEMQWTTQCKPDTTCVQBERELHQNVCFFSVFA
QY	61	KYHFRFRVAVKVOYVDLHLYHLYNEMQWTTQCKPDTTCVQBERELHQNVCFFSVFA	120	KYHFRFRVAVKVOYVDLHLYHLYNEMQWTTQCKPDTTCVQBERELHQNVCFFSVFA
Db	121	VPWFPEQYKIINKSCSSD 137		
QY	121	VPWFPEQYKIINKSCSSD 137		
Db	121	VPWFPEQYKIINKSCSSD 137		
QY	121	VPWFPEQYKIINKSCSSD 137		

```
RESULT 2
US-09-941-314-4
; Sequence 4, Application US/09941314
; Patent No. US20020142396A1
; GENERAL INFORMATION:
; APPLICANT: ZymoGenetics, Inc.
; TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to
; TITLE OF INVENTION: Inhibit Cancer Procoagulant Protein
; FILE REFERENCE: 00-81PC
; CURRENT APPLICATION NUMBER: US/09/941,314
; CURRENT FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/230,230
; PRIOR FILING DATE: 2001-09-01
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-941-314-4

Query Match      87.0%; Score 640; DB 10; Length 117;
Best Local Similarity 100.0%; Pred. No. 1.6e-62;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 21 LPYARKKTFLSVHEWAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKVRQVTDHL 80
Db 1 LPYARKKTFLSVHEWAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKVRQVTDHL 60

Cy 81 EYHNLNEMQWTTCKPRTTNCVPOREHLKQVNCFFSVFVAVPWFEOYKILNKSCSSD 137
Db 61 EYHNLNEMQWTTCKPRTTNCVPOREHLKQVNCFFSVFVAVPWFEOYKILNKSCSSD 117

RESULT 3
US-09-941-314-3
; Sequence 3, Application US/09941314
; Patent No. US20020142396A1
; GENERAL INFORMATION:
; APPLICANT: ZymoGenetics, Inc.
; TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to
; TITLE OF INVENTION: Inhibit Cancer Procoagulant Protein
; FILE REFERENCE: 00-81PC
; CURRENT APPLICATION NUMBER: US/09/941,314
; CURRENT FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/230,230
; PRIOR FILING DATE: 2001-09-01
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-941-314-3

Query Match      85.5%; Score 629; DB 10; Length 115;
Best Local Similarity 100.0%; Pred. No. 2.6e-61;
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 23 YQARKTFLSVHEWAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKVRQVTDHLEY 82
Db 1 YQARKTFLSVHEWAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKVRQVTDHLEY 60

Cy 83 HLANEMQWTTCKPRTTNCVPOREHLKQVNCFFSVFVAVPWFEOYKILNKSCSSD 137
Db 61 HLANEMQWTTCKPRTTNCVPOREHLKQVNCFFSVFVAVPWFEOYKILNKSCSSD 115

RESULT 4
US-09-941-314-15
; Sequence 15, Application US/09941314
; Patent No. US20020142396A1
; GENERAL INFORMATION:
```

```
; APPLICANT: ZymoGenetics, Inc.
; TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to
; TITLE OF INVENTION: Inhibit Cancer Procoagulant Protein
; FILE REFERENCE: 00-81PC
; CURRENT APPLICATION NUMBER: US/09/941,314
; CURRENT FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/230,230
; PRIOR FILING DATE: 2001-09-01
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-941-314-15

Query Match      60.6%; Score 446; DB 10; Length 80;
Best Local Similarity 100.0%; Pred. No. 2e-41;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 53 QYNKESDDKXHFRIFRVLKVRQVTDHLEYHNLNEMQWTTCKPRTTNCVPOREHLKQV 112
Db 1 QYNKESDDKXHFRIFRVLKVRQVTDHLEYHNLNEMQWTTCKPRTTNCVPOREHLKQV 60

Cy 113 NCFPSVAVPWFEOYKILNK 132
Db 61 NCFPSVAVPWFEOYKILNK 80

RESULT 5
US-09-941-314-16
; Sequence 16, Application US/09941314
; Patent No. US20020142396A1
; GENERAL INFORMATION:
; APPLICANT: ZymoGenetics, Inc.
; TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to
; TITLE OF INVENTION: Inhibit Cancer Procoagulant Protein
; FILE REFERENCE: 00-81PC
; CURRENT APPLICATION NUMBER: US/09/941,314
; CURRENT FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/230,230
; PRIOR FILING DATE: 2001-09-01
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-941-314-16

Query Match      45.4%; Score 334; DB 10; Length 59;
Best Local Similarity 100.0%; Pred. No. 2.7e-29;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 74 RQVTDHLEYHNLNEMQWTTCKPRTTNCVPOREHLKQVNCFFSVFVAVPWFEOYKILNK 132
Db 1 RQVTDHLEYHNLNEMQWTTCKPRTTNCVPOREHLKQVNCFFSVFVAVPWFEOYKILNK 59

RESULT 6
US-09-941-314-14
; Sequence 14, Application US/09941314
; Patent No. US20020142396A1
; GENERAL INFORMATION:
; APPLICANT: ZymoGenetics, Inc.
; TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to
; TITLE OF INVENTION: Inhibit Cancer Procoagulant Protein
; FILE REFERENCE: 00-81PC
; CURRENT APPLICATION NUMBER: US/09/941,314
; CURRENT FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/230,230
; PRIOR FILING DATE: 2001-09-01
; NUMBER OF SEQ ID NOS: 19
```

SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 14  
LENGTH: 52  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-941-314-14

Query Match 39.1%; Score 288; DB 10; Length 52;  
Best Local Similarity 100.0%; Pred. No. 2.6e-24;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 KESDDKHFRIFRVLKQROVTDHLEHNLVNMQWTTCKPRTNCPQERE 107  
DB 1 KESDDKHFRIFRVLKQROVTDHLEHNLVNMQWTTCKPRTNCPQERE 52

RESULT 7  
US-09-941-314-17  
Sequence 17, Application US/09941314  
Patent No. US20020142396A1  
GENERAL INFORMATION:  
APPLICANT: ZymoGenetics, Inc.  
TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to  
FILE REFERENCE: 00-81PC  
CURRENT APPLICATION NUMBER: US/09/941,314  
CURRENT FILING DATE: 2001-08-29  
PRIOR APPLICATION NUMBER: 60/230,230  
PRIOR FILING DATE: 2001-09-01  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 17  
LENGTH: 48  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-941-314-17

Query Match 37.1%; Score 273; DB 10; Length 48;  
Best Local Similarity 100.0%; Pred. No. 1.1e-22;  
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 NVEWQWTTCKPRTNCPQEREHAKQVNCFFSVFANPWFQYKILNK 132  
DB 1 NVEWQWTTCKPRTNCPQEREHAKQVNCFFSVFANPWFQYKILNK 48

RESULT 8  
US-09-941-314-13  
Sequence 13, Application US/09941314  
Patent No. US20020142396A1  
GENERAL INFORMATION:  
APPLICANT: ZymoGenetics, Inc.  
TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to  
FILE REFERENCE: 00-81PC  
CURRENT APPLICATION NUMBER: US/09/941,314  
CURRENT FILING DATE: 2001-08-29  
PRIOR APPLICATION NUMBER: 60/230,230  
PRIOR FILING DATE: 2001-09-01  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13  
LENGTH: 49  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-941-314-13

Query Match 37.0%; Score 272; DB 10; Length 49;  
Best Local Similarity 100.0%; Pred. No. 1.4e-22;  
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DQYNKESDDKHFRIFRVLKQROVTDHLEHNLVNMQWTTCKPRTN 100

DB 1 DQYNKESDDKHFRIFRVLKQROVTDHLEHNLVNMQWTTCKPRTN 49

RESULT 9  
US-09-864-761-34822  
Sequence 34822, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 34822  
LENGTH: 50  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL109954.10  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.7  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.2  
OTHER INFORMATION: EST HUMAN HIT: A1200857.1, EVALUATE 5.00e-23  
OTHER INFORMATION: SWISSPROT HIT: O60676, EVALUATE 1.00e-01

US-09-864-761-34822

Query Match 35.6%; Score 262; DB 9; Length 50;  
Best Local Similarity 100.0%; Pred. No. 1.8e-21;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26 RKKTFLSVHEVMAVENYAKDSLQWITDQYNKESDDKHFRIFRVLKVRQ 75  
Db 1 RKKTFLSVHEVMAVENYAKDSLQWITDQYNKESDDKHFRIFRVLKVRQ 50

RESULT 10  
US-09-864-761-48936

; Sequence 48936, Application US/09864761  
; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharon G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEOTIC ACID PROBES USEFUL FOR

; FILE REFERENCE: Aecm1ca-X-1

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US/09/864,761

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: GB 24263,6

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

; PRIOR FILING DATE: 2000-06-30

; PRIOR APPLICATION NUMBER: US 09/774,203

; PRIOR FILING DATE: 2001-01-29

; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

; SEQ ID NO 48936

; LENGTH: 50

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: MAP TO AI096677.18

; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.96

; OTHER INFORMATION: EST HUMAN HIT: AI200857.1, EVALUATE 5.00e-23

; OTHER INFORMATION: SWISSPROT HIT: O60676, EVALUATE 1.00e-01

; US-09-864-761-48936

Query Match 35.6%; Score 262; DB 9; Length 50;  
Best Local Similarity 100.0%; Pred. No. 1.8e-21;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26 RKKTFLSVHEVMAVENYAKDSLQWITDQYNKESDDKHFRIFRVLKVRQ 75  
Db 1 RKKTFLSVHEVMAVENYAKDSLQWITDQYNKESDDKHFRIFRVLKVRQ 50

RESULT 11  
US-09-941-314-10

; Sequence 10, Application US/09941314  
; Patent No. US20020142396A1

; GENERAL INFORMATION:

; APPLICANT: ZymoGenetics, Inc.

; TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to

; FILE REFERENCE: 00-81PC

; CURRENT FILING DATE: 2001-09-01

; PRIOR APPLICATION NUMBER: US/09/941,314

; PRIOR FILING DATE: 2001-08-29

; PRIOR APPLICATION NUMBER: 60/230,230

; PRIOR FILING DATE: 2001-09-01

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 10

; LENGTH: 46

; TYPE: PRT

; ORGANISM: Homo sapiens

; US-09-941-314-10

Query Match 34.5%; Score 254; DB 10; Length 46;  
Best Local Similarity 100.0%; Pred. No. 1.2e-20;  
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 NKESDDKHFRIFRVLKVRQVTDHLEHNLVEMQWTCCKPPTTN 100  
Db 1 NKESDDKHFRIFRVLKVRQVTDHLEHNLVEMQWTCCKPPTTN 46

RESULT 12  
US-09-941-314-12

; Sequence 12, Application US/09941314  
; Patent No. US20020142396A1

; GENERAL INFORMATION:

; APPLICANT: ZymoGenetics, Inc.

; TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to

; FILE REFERENCE: 00-81PC

; CURRENT FILING DATE: 2001-08-29

; PRIOR APPLICATION NUMBER: US/09/941,314

; PRIOR FILING DATE: 2001-09-01

; PRIOR APPLICATION NUMBER: 60/230,230

; PRIOR FILING DATE: 2001-09-01

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 12

; LENGTH: 33

; TYPE: PRT

; ORGANISM: Homo sapiens

; US-09-941-314-12

Query Match 25.7%; Score 189; DB 10; Length 33;  
Best Local Similarity 100.0%; Pred. No. 1.1e-13;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 75 QVTDHLEHNLVEMQWTCCKPPTTNQVPOERE 107  
Db 1 QVTDHLEHNLVEMQWTCCKPPTTNQVPOERE 33

RESULT 13  
US-09-941-314-6  
; Sequence 6, Application US/09941314

Patent No. US20020142396A1  
GENERAL INFORMATION:  
APPLICANT: ZymoGenetics, Inc.  
TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to  
FILE REFERENCE: 00-81PC  
CURRENT APPLICATION NUMBER: US/09/941,314  
CURRENT FILING DATE: 2001-08-29  
PRIOR APPLICATION NUMBER: 60/230,230  
PRIOR FILING DATE: 2001-09-01  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 6  
LENGTH: 36  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-941-314-6

Query Match 25.7%; Score 189; DB 10; Length 36;  
Best Local Similarity 100.0%; Pred. No. 1.2e-13;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 RKKTPLSVHEVMAVENYAKDSLQWITDOYNKESDDK 61  
DB 1 RKKTPLSVHEVMAVENYAKDSLQWITDOYNKESDDK 36

RESULT 14  
US-09-941-314-8  
Sequence 8, Application US/09941314  
Patent No. US20020142396A1  
GENERAL INFORMATION:  
APPLICANT: ZymoGenetics, Inc.  
TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to  
FILE REFERENCE: 00-81PC  
CURRENT APPLICATION NUMBER: US/09/941,314  
CURRENT FILING DATE: 2001-08-29  
PRIOR APPLICATION NUMBER: 60/230,230  
PRIOR FILING DATE: 2001-09-01  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 8  
LENGTH: 35  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-941-314-8

Query Match 25.4%; Score 187; DB 10; Length 35;  
Best Local Similarity 100.0%; Pred. No. 2e-13;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 KDSLQWITDOYNKESDDKHYFRIFRYLKVQROVTD 78  
DB 1 KDSLQWITDOYNKESDDKHYFRIFRYLKVQROVTD 35

RESULT 15  
US-10-376-564-46  
Sequence 46, Application US/10376564  
Publication No. US20030180302A1  
GENERAL INFORMATION:  
APPLICANT: Wolf, Eckhard  
APPLICANT: Werner, Sabine  
APPLICANT: Halle, Jörn-Peter  
APPLICANT: Regenbogen, Johannes  
APPLICANT: Goppelt, Andreas  
TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for  
TITLE OF INVENTION: the Diagnosis or Treatment of Skin Disorders and Wound  
TITLE OF INVENTION: Healing and for the Identification of Pharmacologically  
FILE REFERENCE: 50125/014003  
CURRENT APPLICATION NUMBER: US/10/376,564

CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 09/886,319  
PRIOR FILING DATE: 2001-06-20  
PRIOR APPLICATION NUMBER: US 60/222,081  
PRIOR FILING DATE: 2000-08-01  
PRIOR APPLICATION NUMBER: DE 100 30 149.5-41  
PRIOR FILING DATE: 2000-06-20  
NUMBER OF SEQ ID NOS: 86  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 46  
LENGTH: 140  
TYPE: PRT  
ORGANISM: Mus musculus  
US-10-376-564-46

Query Match 25.3%; Score 186.5; DB 12; Length 140;  
Best Local Similarity 30.9%; Pred. No. 1.3e-12;  
Matches 43; Conservative 30; Mismatches 63; Indels 3; Gaps 2;

QY 1 MASPPQALQLLAILLTMAIPYQARKFTLSVHEVMAVENYAKDSLQWITDOYNKESDD 60  
DB 1 MASPLRSLLFLILAVLAVAWATPKQGPRLMGABEADANEGBVRRALDFAVSEYKNSND 60  
QY 61 KYHFRIFRYLKVQROVTDHLEVLNVMQWTCQKDET--TNC-VPGERELHROVNGEFS 117  
DB 61 AYHSRAIQVVRARQQLVAGVNYFLDVMGRTTCTKSQTNLIDCPHIDQPHLMRALCSFQ 120  
QY 118 VFAVPWFQYKILNKSCSS 136  
DB 121 IYSPWKQTHSLTKFSCKN 139

Search completed: January 21, 2004, 12:17:39  
Job time : 49.6129 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 12:24:32 ; Search time 21 Seconds  
(without alignments)  
627.386 Million cell updates/sec

Title: US-09-941-314-2  
Perfect score: 137  
Sequence: 1 MAEPWQALQLLAAILTLTA.....VFAVPWFQYKILNKSCSSD 137

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 9616682 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Database : PIR 76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	5.8	707	2 AP2277	serine/threonine k
2	8	5.8	710	2 T26742	hypothetical prote
3	8	5.8	1966	2 T32552	hypothetical prote
4	8	5.8	2848	2 T32550	hypothetical prote
5	7	5.1	75	2 S70178	hypothetical prote
6	7	5.1	89	2 C86497	hypothetical prote
7	7	5.1	105	2 H72125	hypothetical prote
8	7	5.1	118	1 NNIVX4	nonstructural prote
9	7	5.1	118	1 NNIVXX	nonstructural prote
10	7	5.1	118	1 NNIVX8	nonstructural prote
11	7	5.1	118	1 NNIVX2	nonstructural prote
12	7	5.1	118	1 NNIVX6	nonstructural prote
13	7	5.1	121	1 NNIV2W	nonstructural prote
14	7	5.1	121	1 E45539	nonstructural prote
15	7	5.1	121	1 NNIVB1	nonstructural prote
16	7	5.1	121	1 NNIVB3	nonstructural prote
17	7	5.1	121	1 NNIVB4	nonstructural prote
18	7	5.1	121	1 NNIV26	nonstructural prote
19	7	5.1	121	1 NNIVB5	nonstructural prote
20	7	5.1	121	1 NNIVB6	nonstructural prote
21	7	5.1	121	1 NNIVB7	nonstructural prote
22	7	5.1	121	1 NNIVB8	nonstructural prote
23	7	5.1	121	1 B55575	nonstructural prote
24	7	5.1	148	2 S33685	probable membrane
25	7	5.1	148	2 A11058	hypothetical prote
26	7	5.1	164	2 B72717	hypothetical prote
27	7	5.1	184	2 D72061	hypothetical prote
28	7	5.1	184	2 B86562	hypothetical prote
29	7	5.1	225	2 C71159	hypothetical prote

30	7	5.1	248	2 C90915	hypothetical prote
31	7	5.1	248	2 B85763	hypothetical prote
32	7	5.1	283	2 B83459	probable potassium
33	7	5.1	288	2 C86687	hypothetical prote
34	7	5.1	296	2 S24386	protein kinase (EC
35	7	5.1	301	2 G83556	hypothetical prote
36	7	5.1	330	2 F64905	probable sugar tra
37	7	5.1	330	2 B90894	probable transport
38	7	5.1	330	2 G85723	probable transport
39	7	5.1	332	1 S41003	protein kinase (EC
40	7	5.1	332	2 AG3026	hypothetical prote
41	7	5.1	332	2 B98258	ribose ABC transpo
42	7	5.1	359	2 T20575	hypothetical prote
43	7	5.1	359	2 B71516	probable Fe-S oxid
44	7	5.1	369	2 F81674	conserved hypochet
45	7	5.1	375	2 F71234	probable Na+/H+-ex
46	7	5.1	386	2 T04914	hypothetical prote
47	7	5.1	401	2 S20035	pulF protein - Kle
48	7	5.1	429	2 AC1163	flagellar hook-ass
49	7	5.1	429	2 AC1522	flagellar hook-ass
50	7	5.1	435	2 T46443	hypothetical prote
51	7	5.1	445	2 JA0048	tubulin beta-1 cha
52	7	5.1	479	2 F70965	hypothetical prote
53	7	5.1	497	2 JB0275	voltage-gated pota
54	7	5.1	506	2 C95217	choline transporte
55	7	5.1	506	2 T44634	hypothetical prote
56	7	5.1	506	2 C98081	hypothetical prote
57	7	5.1	676	2 G69154	conserved hypochet
58	7	5.1	700	2 B84131	transcription anti
59	7	5.1	702	2 T21148	hypothetical prote
60	7	5.1	878	2 T08559	protein kinase hom
61	7	5.1	2287	2 AB2494	hypothetical prote
62	7	5.1	3396	1 A42551	genome polypeptid
63	6	4.4	36	2 B82332	hypothetical prote
64	6	4.4	54	2 F34284	H+-transporting tw
65	6	4.4	61	2 AC2556	hypothetical prote
66	6	4.4	65	2 B75298	hypothetical prote
67	6	4.4	73	2 C95909	conserved hypochet
68	6	4.4	77	2 B97156	exonuclease VII sm
69	6	4.4	79	2 A95090	KH domain protein
70	6	4.4	79	2 F75343	hypothetical prote
71	6	4.4	79	2 C97957	conserved hypochet
72	6	4.4	80	2 A33286	ubiquitinol-cytochro
73	6	4.4	92	2 AG3634	norf (imported) -
74	6	4.4	93	2 S14314	neutrophil cationi
75	6	4.4	93	2 S21169	neutrophil cationi

#### ALIGNMENTS

RESULT 1  
AP2277  
serine/threonine kinase (imported) - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. Strain PCC 7120 is a synonym of Anabaena sp. Strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S  
DNA Ref: 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anai  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AF2277  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-707 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA075472.1; PID:G17132907; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: all3773

Query Match 5.8% Score 8; DB 2; Length 707;

Best Local Similarity 100.0%; Pred. No. 3.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 LLLAIIIL 17  
Db 303 LLLAIIIL 310

## RESULT 2

T26742  
hypothetical protein Y39A1A.22 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C/Accession: T26742

R/Wall, M.  
submitted to the EMBL Data Library, September 1998

A/Reference number: Z20257

A/Accession: T26742

A/Molecule type: DNA

A/Residues: 1-710 <JUH>

A/Cross-references: EMBL:AL031633, PIDN:CAA21031.1, GSPDB:GN00021, CESP:Y39A1A.22

A/Experimental source: clone Y39A1A

C/Genetics:

A/Gene: CESP:Y39A1A.22

A/Map position: 3

A/Introns: 212/3

Query Match  
Best Local Similarity 100.0%; Pred. No. 3.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 LLLAIIIL 17  
Db 238 LLLAIIIL 245

## RESULT 3

T32552  
hypothetical protein F33D4.2c - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 01-Dec-2000

C/Accession: T32552

R/Johnson, D.; Steillyes, L.  
submitted to the EMBL Data Library, December 1997

A/Description: The sequence of C. elegans cosmid F33D4.

A/Reference number: Z21190

A/Accession: T32552

A/Molecule type: DNA

A/Residues: 1-1966 <JUH>

A/Cross-references: EMBL:AF036702, PIDN:AA88380.1, GSPDB:GN00022, CESP:F33D4.2c

A/Experimental source: strain Bristol N2; clone F33D4

A/Accession: T32551

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1844, 'G', 1917-1918, 'AF', 1921-1922 <JUH>

A/Cross-references: EMBL:AF036702, PIDN:AA88375.1, GSPDB:GN00022, CESP:F33D4.2b

A/Experimental source: strain Bristol N2; clone F33D4

C/Genetics:

A/Gene: CESP:F33D4.2c; CESP:F33D4.2b

A/Map position: 4

A/Introns: 48/2; 117/3; 167/3; 263/3; 316/3; 362/1; 378/2; 396/1; 436/3; 509/3; 658/3; 7

## Query Match

Best Local Similarity 100.0%; Pred. No. 9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLLAIIIL 16  
Db 1093 QLLAIIIL 1100

RESULT 4  
T32550  
hypothetical protein F33D4.2a - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000

C/Accession: T32550

R/Johnson, D.; Steillyes, L.

submitted to the EMBL Data Library, December 1997

A/Description: The sequence of C. elegans cosmid F33D4.

A/Reference number: Z21190

A/Accession: T32550

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-2848 <JUH>

A/Cross-references: EMBL:AF036702, PIDN:AA88374.1, GSPDB:GN00022, CESP:F33D4.2a

A/Experimental source: strain Bristol N2; clone F33D4

C/Genetics:

A/Gene: CESP:F33D4.2a

A/Map position: 4

A/Introns: 48/2; 117/3; 167/3; 263/3; 316/3; 362/1; 378/2; 396/1; 436/3; 509/3; 658/3; 7

C/Superfamily: inositol-trisphosphate receptor

Query Match  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLLAIIIL 16  
Db 1093 QLLAIIIL 1100

## RESULT 5

S70178  
hypothetical protein X - Rhodobacter sphaeroides

C/Species: Rhodobacter sphaeroides

C/Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 08-Oct-1999

C/Accession: S70178; S49209

R/Ward, M.J.; Bell, A.W.; Hamblin, P.A.; Packer, H.L.; Amtege, J.P.

Mol. Microbiol. 17, 357-366, 1995

A/Title: Identification of a chemotaxis operon with two che Y genes in Rhodobacter sphaer

A/Reference number: S70178; MUID:96079285; PMID:7494484

A/Accession: S70178

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-75 <MAR>

A/Cross-references: EMBL:X80027, NID:9510668, PIDN:CAA56328.1; PID:9510669

A/Experimental source: strain WS8-N

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994

Query Match  
Best Local Similarity 100.0%; Pred. No. 6.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ALQIIIL 13  
Db 26 ALQIIIL 32

## RESULT 6

C86497  
hypothetical protein CP10050 [imported] - Chlamydia pneumoniae (strain J138)

C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001

C/Accession: C86497

R/Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise

Nucleic Acids Res. 28, 2311-2314, 2000

A/Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.

A/Reference number: A86491, MUID:20330349; PMID:10871362

A/Accession: C86497

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-89 <STO>

A:Cross-references: GB:BA000008; NID:G8978424; PIDN:BA98261.1; GSPDB:GN00142  
A:Experimental source: strain J138  
C:Genetics:  
A:Gene: CPJ0050

Query Match 5.1%; Score 7; DB 2; Length 89;  
Best Local Similarity 100.0%; Pred. No. 7.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIIIL 16  
|||||  
Db 60 LLLAIIIL 66

RESULT 7  
H72125  
hypothetical protein - Chlamydia pneumoniae (strain CWL029)  
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
C/Accession: H72125

R:Nakajima, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606; PMID:10192388  
A:Accession: H72125  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-105 <NAN>  
A:Cross-references: GB:AE001590; GB:AE001363; NID:G4376299; PIDN:AD18203.1; PID:G437630  
C:Experimental source: strain CWL029  
C:Genetics:  
A:Gene: CPN0050

Query Match 5.1%; Score 7; DB 2; Length 105;  
Best Local Similarity 100.0%; Pred. No. 8.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIIIL 16  
|||||  
Db 60 LLLAIIIL 66

RESULT 8  
MN1VX4  
nonstructural protein NS2 - influenza A virus (strain A/Duck/England/56 [H1N6]) (fragment)

C:Species: influenza A virus  
C>Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 16-Jul-1999  
C/Accession: D27846  
R:Nakajima, K.; Nobusawa, E.; Ogawa, T.; Nakajima, S.  
Virology 158, 465-468, 1987  
A:Title: Genetic divergence of the NS genes of avian influenza viruses.  
A:Reference number: A94361; MUID:87236215; PMID:2954302  
A:Accession: D27846  
A:Molecule type: genomic RNA  
A:Residues: 1-118 <NAN>  
A:Cross-references: GB:M16563; NID:G324786; PIDN:AAA43511.1; PID:G324788  
C:Genetics:  
A:Gene: NS2  
A:Map position: segment 8  
A:Introns: 7/3  
C:Superfamily: influenza virus nonstructural protein NS2  
C:Keywords: alternative splicing; nonstructural protein

Query Match 5.1%; Score 7; DB 1; Length 118;  
Best Local Similarity 100.0%; Pred. No. 9.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQILL 12  
|||||  
Db 98 QALQILL 104

RESULT 9

MN1VXX  
nonstructural protein NS2 - influenza A virus (strain A/Wyrmh/Haneda-Thai/76 [H3N1]) (fragment)

C:Species: influenza A virus  
C>Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 16-Jul-1999  
C/Accession: A30086  
R:Nakajima, K.; Nobusawa, E.; Ogawa, T.; Nakajima, S.  
Virology 158, 465-468, 1987  
A:Title: Genetic divergence of the NS genes of avian influenza viruses.  
A:Reference number: A94361; MUID:87236215; PMID:2954302  
A:Accession: A30086  
A:Molecule type: genomic RNA  
A:Residues: 1-118 <NAN>  
A:Cross-references: GB:M17070; NID:G324854; PIDN:AAA43549.1; PID:G324856  
C:Genetics:  
A:Gene: NS2  
A:Map position: segment 8  
A:Introns: 7/3  
C:Superfamily: influenza virus nonstructural protein NS2  
C:Keywords: alternative splicing; nonstructural protein

Query Match 5.1%; Score 7; DB 1; Length 118;  
Best Local Similarity 100.0%; Pred. No. 9.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQILL 12  
|||||  
Db 98 QALQILL 104

RESULT 10  
MN1VX8  
nonstructural protein NS2 - influenza A virus (strain A/Duck/Ukraine/63 [H3N8]) (fragment)

C:Species: influenza A virus  
C>Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 16-Jul-1999  
C/Accession: H27846  
R:Nakajima, K.; Nobusawa, E.; Ogawa, T.; Nakajima, S.  
Virology 158, 465-468, 1987  
A:Title: Genetic divergence of the NS genes of avian influenza viruses.  
A:Reference number: A94361; MUID:87236215; PMID:2954302  
A:Accession: H27846  
A:Molecule type: genomic RNA  
A:Residues: 1-118 <NAN>  
A:Cross-references: GB:M16565; NID:G324789; PIDN:AAA43513.1; PID:G324791  
C:Genetics:  
A:Gene: NS2  
A:Map position: segment 8  
A:Introns: 7/3  
C:Superfamily: influenza virus nonstructural protein NS2  
C:Keywords: alternative splicing; nonstructural protein

Query Match 5.1%; Score 7; DB 1; Length 118;  
Best Local Similarity 100.0%; Pred. No. 9.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQILL 12  
|||||  
Db 98 QALQILL 104

RESULT 11

MN1VX2  
nonstructural protein NS2 - influenza A virus (strain A/Chicken/Japan/24 [H7N7]) (fragment)

C:Species: influenza A virus  
C>Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 16-Jul-1999  
C/Accession: B27846  
R:Nakajima, K.; Nobusawa, E.; Ogawa, T.; Nakajima, S.  
Virology 158, 465-468, 1987  
A:Title: Genetic divergence of the NS genes of avian influenza viruses.  
A:Reference number: A94361; MUID:87236215; PMID:2954302  
A:Accession: B27846  
A:Molecule type: genomic RNA  
A:Residues: 1-118 <NAN>  
A:Cross-references: GB:M16561; NID:G324777; PIDN:AAA43505.1; PID:G324779

Query Match 5.1%; Score 7; DB 1; Length 118;  
Best Local Similarity 100.0%; Pred. No. 9.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

C:Genetics:

A:Gene: NS2

A:Map position: segment 8

A:Insertions: 7/3

C:Superfamily: Influenza virus nonstructural protein NS2

C:Keywords: alternative splicing; nonstructural protein

Query Match 5.1%; Score 7; DB 1; Length 118;

Best Local Similarity 100.0%; Pred. No. 9.4; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLL 12

Db 98 QALQLL 104

RESULT 12

N1VX6

nonstructural protein NS2 - influenza A virus (strain A/Tern/South Africa/61 [H5N3]) (F2

C:Species: Influenza A virus

C&gt;Date: 30-Sep-1999 #sequence\_revision 30-Sep-1999 #text\_change 16-Jul-1999

C:Accession: F27846

R:Nakajima, K.; Nobusawa, E.; Ogawa, T.; Nakajima, S.

Virology 158, 465-468, 1987

A:Title: Genetic divergence of the NS genes of avian influenza viruses.

A:Reference number: A94361; MUID:87236215; PMID:2954302

A:Accession: F27846

A:Molecule type: genomic RNA

A:Residues: 1-118 &lt;NAK&gt;

A:Cross-references: GB:M16564; NID:G324875; PIDN:AAA43573.1; PID:G324877

C:Genetics:

A:Gene: NS2

A:Map position: segment 8

A:Insertions: 7/3

C:Superfamily: Influenza virus nonstructural protein NS2

C:Keywords: alternative splicing; nonstructural protein

Query Match 5.1%; Score 7; DB 1; Length 118;

Best Local Similarity 100.0%; Pred. No. 9.4; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLL 12

Db 98 QALQLL 104

RESULT 13

N1V2W

nonstructural protein NS2 - influenza A virus (strain A/FW/1/50 [H1N1])

C:Species: Influenza A virus

C&gt;Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 24-Sep-1999

C:Accession: C92991; A04095

R:Kyteal, M.; Buonagurio, D.; Young, J.F.; Palase, P.

J. Virology 45, 547-554, 1983

A:Title: Sequential mutations in the NS genes of influenza virus field strains.

A:Reference number: A92991; MUID:83164298; PMID:6834468

A:Accession: C92991

A:Molecule type: genomic RNA

A:Residues: 1-121 &lt;KRY&gt;

A:Cross-references: GB:K00577; NID:G324808; PIDN:AAA43521.1; PID:G324809

C:Genetics:

A:Map position: segment 8

C:Superfamily: Influenza virus nonstructural protein NS2

C:Keywords: alternative splicing

Query Match 5.1%; Score 7; DB 1; Length 121;

Best Local Similarity 100.0%; Pred. No. 9.6; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLL 12

Db 101 QALQLL 107

RESULT 14

E45539 nonstructural protein NS2 - influenza A virus (strain A/chicken/Brescia/1902 [H7N7])

C:Species: Influenza A virus

C&gt;Date: 24-Feb-1994 #sequence\_revision 24-Feb-1994 #text\_change 26-May-1994

C:Accession: E45539

R:Klimov, A.; Prosch, S.; Schaffer, J.; Bucher, D.

Arch. Virology 122, 143-161, 1992

A:Title: Subtype H7 influenza viruses: comparative antigenic and molecular analysis of

A:Reference number: A45539; MUID:92109567; PMID:1350908

A:Accession: E45539

A:Molecule type: genomic RNA

A:Residues: 1-121 &lt;KLI&gt;

A:Note: sequence extracted from NCBI backbone (NCBI:74250, NCBI:74255)

C:Genetics:

A:Map position: segment 8

C:Superfamily: Influenza virus nonstructural protein NS2

C:Keywords: alternative splicing

Query Match 5.1%; Score 7; DB 1; Length 121;

Best Local Similarity 100.0%; Pred. No. 9.6; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLL 12

Db 101 QALQLL 107

RESULT 15

N1V1

nonstructural protein NS2 - influenza A virus (strains A/pintail/Alberta/268/78 and A/ma)

C:Species: Influenza A virus

C&gt;Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 19-Oct-1995

C:Accession: B32662; D32662

R:Treanor, J.J.; Snyder, M.H.; London, W.T.; Murphy, B.R.

Virology 171, 1-9, 1989

A:Title: The B allele of the NS gene of avian influenza viruses, but not the A allele, at

A:Reference number: A32662; MUID:89299445; PMID:255836

A:Accession: B32662

A:Molecule type: genomic RNA

A:Residues: 1-121 &lt;TR&gt;

A:Experimental source: strain A/pintail/Alberta/268/78

A:Accession: D32662

C:Genetics:

A:Gene: NS2

A:Map position: segment 8

A:Insertions: 10/3

C:Superfamily: Influenza virus nonstructural protein NS2

C:Keywords: alternative splicing; nonstructural protein

Query Match 5.1%; Score 7; DB 1; Length 121;

Best Local Similarity 100.0%; Pred. No. 9.6; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLL 12

Db 101 QALQLL 107

RESULT 16

N1V3

nonstructural protein NS2 - influenza A virus (strain A/mallard/New York/6874/78)

C:Species: Influenza A virus

C&gt;Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 08-Apr-1994

C:Accession: F32662

R:Treanor, J.J.; Snyder, M.H.; London, W.T.; Murphy, B.R.

Virology 171, 1-9, 1989

A:Title: The B allele of the NS gene of avian influenza viruses, but not the A allele, at

A:Reference number: A32662; MUID:89299445; PMID:255836

A/Accession: F32662  
A/Molecule type: genomic RNA  
A/Residues: 1-121 <TR>  
C/Genetics:  
A/Gene: NS2  
A/Map position: segment 8  
A/Intons: 10/3  
C/Superfamily: influenza virus nonstructural protein NS2  
C/Keywords: alternative splicing; nonstructural protein

Query Match  
Best local Similarity 100.0%; Pred. No. 9.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQALL 12  
DB 101 QALQALL 107

RESULT 17  
MNIVB4  
nonstructural protein NS2 - influenza A virus (strain A/pintail/Alberta/119/79)  
C/Species: influenza A virus  
C/Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 08-Apr-1994  
C/Accession: H32662  
R/Treanor, J.J.; Snyder, M.H.; London, W.T.; Murphy, B.R.  
Virology 171, 1-9, 1989  
A/Title: The B allele of the NS gene of avian influenza viruses, but not the A allele,  
A/Reference number: A32662; PMID:89299445; PMID:2525836  
A/Accession: H32662  
A/Molecule type: genomic RNA  
A/Residues: 1-121 <TR>  
C/Genetics:  
A/Gene: NS2  
A/Map position: segment 8  
A/Intons: 10/3  
C/Superfamily: influenza virus nonstructural protein NS2  
C/Keywords: alternative splicing; nonstructural protein

Query Match  
Best local Similarity 100.0%; Pred. No. 9.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQALL 12  
DB 101 QALQALL 107

RESULT 18  
MNIV26  
nonstructural protein NS2 - influenza A virus (strains A/duck/Alberta/60/76 and A/turkey  
C/Species: influenza A virus  
C/Date: 14-Nov-1983 #sequence\_revision 14-Nov-1983 #text\_change 16-Jul-1999  
C/Accession: A04098; B27529  
R/Beetz, M.; Zazra, J.J.; Elliott, R.M.; Young, J.F.; Palese, P.  
Virology 113, 397-402, 1981  
A/Title: Nucleotide sequence of the influenza A/duck/Alberta/60/76 virus NS RNA: conserv  
A/Reference number: A04092; PMID:81276929; PMID:6927848  
A/Accession: A04098  
A/Molecule type: genomic RNA  
A/Residues: 1-121 <BMR>  
A/Cross-references: GB:002105; GB:M17071; NID:G324783; PIDN:AAA43508.1; PID:G324784  
A/Experimental source: strain A/duck/Alberta/60/76  
R/Norton, G.P.; Tanaka, T.; Tobita, K.; Nakada, S.; Buonagurio, D.A.; Greenspan, D.; Ky  
Virology 156, 204-213, 1987  
A/Title: Infectious influenza A and B virus variants with long carboxyl terminal deletio  
A/Reference number: A27529; PMID:87122162; PMID:3811235  
A/Accession: B27529  
A/Molecule type: genomic RNA  
A/Residues: 1-121 <NOR>  
A/Experimental source: strain A/turkey/Oregon/71 (H7N5)  
C/Genetics:  
A/Map position: segment 8

C/Superfamily: influenza virus nonstructural protein NS2  
C/Keywords: alternative splicing; nonstructural protein

Query Match  
Best local Similarity 100.0%; Pred. No. 9.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQALL 12  
DB 101 QALQALL 107

RESULT 19  
MNIVB5  
nonstructural protein NS2 - influenza A virus (strain A/mallard/Alberta/86/76)  
C/Species: influenza A virus  
C/Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 08-Apr-1994  
C/Accession: B32663  
R/Treanor, J.J.; Snyder, M.H.; London, W.T.; Murphy, B.R.  
Virology 171, 1-9, 1989  
A/Title: The B allele of the NS gene of avian influenza viruses, but not the A allele,  
A/Reference number: A32662; PMID:89299445; PMID:2525836  
A/Accession: B32663  
A/Molecule type: genomic RNA  
A/Residues: 1-121 <TR>  
C/Genetics:  
A/Gene: NS2  
A/Map position: segment 8  
A/Intons: 10/3  
C/Superfamily: influenza virus nonstructural protein NS2  
C/Keywords: alternative splicing; nonstructural protein

Query Match  
Best local Similarity 100.0%; Pred. No. 9.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQALL 12  
DB 101 QALQALL 107

RESULT 20  
MNIVB6  
nonstructural protein NS2 - influenza A virus (strain A/mallard/Alberta/827/78)  
C/Species: influenza A virus  
C/Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 08-Apr-1994  
C/Accession: D32663  
R/Treanor, J.J.; Snyder, M.H.; London, W.T.; Murphy, B.R.  
Virology 171, 1-9, 1989  
A/Title: The B allele of the NS gene of avian influenza viruses, but not the A allele,  
A/Reference number: A32662; PMID:89299445; PMID:2525836  
A/Accession: D32663  
A/Molecule type: genomic RNA  
A/Residues: 1-121 <TR>  
C/Genetics:  
A/Gene: NS2  
A/Map position: segment 8  
A/Intons: 10/3  
C/Superfamily: influenza virus nonstructural protein NS2  
C/Keywords: alternative splicing; nonstructural protein

Query Match  
Best local Similarity 100.0%; Pred. No. 9.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQALL 12  
DB 101 QALQALL 107

RESULT 21  
MNIVB7  
nonstructural protein NS2 - influenza A virus (strain A/pintail/Alberta/121/79)

C/Species: Influenza A virus  
C/Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 08-Apr-1994  
C/Accession: F32663  
R/Treanor, J.J.; Snyder, M.H.; London, W.T.; Murphy, B.R.  
Virology 171, 1-9, 1989  
A/Title: The B allele of the NS gene of avian influenza viruses, but not the A allele, a  
A/Reference number: A32662; MUID:89299445; PMID:2525836  
A/Accession: F32663  
A/Molecule type: genomic RNA  
A/Residues: 1-121 <TRB>  
C/Genetics:  
A/Gene: NS2  
A/Map position: segment 8  
A/Intons: 10/3  
C/Superfamily: influenza virus nonstructural protein NS2  
C/Keywords: alternative splicing; nonstructural protein

Query Match  
5.1%; Score 7; DB 1; Length 121;  
Best Local Similarity 100.0%; Pred. No. 9.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLL 12  
| | | | |  
Db 101 QALQLL 107

RESULT 22  
MNIVB8  
nonstructural protein NS2 - influenza A virus (strain A/pintail/Alberta/358/79)  
C/Species: Influenza A virus  
C/Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 08-Apr-1994  
C/Accession: H32663  
R/Treanor, J.J.; Snyder, M.H.; London, W.T.; Murphy, B.R.  
Virology 171, 1-9, 1989  
A/Title: The B allele of the NS gene of avian influenza viruses, but not the A allele, a  
A/Reference number: A32662; MUID:89299445; PMID:2525836  
A/Accession: H32663  
A/Molecule type: genomic RNA  
A/Residues: 1-121 <TRB>  
C/Genetics:  
A/Gene: NS2  
A/Map position: segment 8  
A/Intons: 10/3  
C/Superfamily: influenza virus nonstructural protein NS2  
C/Keywords: alternative splicing; nonstructural protein

Query Match  
5.1%; Score 7; DB 1; Length 121;  
Best Local Similarity 100.0%; Pred. No. 9.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLL 12  
| | | | |  
Db 101 QALQLL 107

RESULT 23  
B45575  
nonstructural protein NS2 - influenza A virus (strain A/Turkey/Wisconsin/68 [H5N9])  
C/Species: Influenza A virus  
C/Date: 22-Apr-1993 #sequence\_revision 22-Apr-1993 #text\_change 22-May-1998  
C/Accession: B45575  
R/Perdue, M.L.  
Virus Res. 23, 223-240, 1992  
A/Title: Naturally occurring NS gene variants in an avian influenza virus isolate.  
A/Reference number: A45575; MUID:92327829; PMID:1320795  
A/Accession: B45575  
A/Molecule type: genomic RNA  
A/Residues: 1-121 <PBR>  
A/Experimental source: A/Turkey/Wisconsin/68, H5N9  
A/Note: sequence inconsistent with the nucleotide translation  
C/Superfamily: influenza virus nonstructural protein NS2  
C/Keywords: alternative splicing; nonstructural protein

Query Match  
5.1%; Score 7; DB 2; Length 121;  
Best Local Similarity 100.0%; Pred. No. 9.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLL 12  
| | | | |  
Db 101 QALQLL 107

RESULT 24  
S33685  
nonstructural protein NS1 - influenza A virus (strain A/WS/33)  
C/Species: Influenza A virus  
C/Variety: strain A/WS/33  
C/Date: 02-Dec-1993 #sequence\_revision 26-May-1995 #text\_change 20-Sep-1999  
C/Accession: S33685; S34221  
R/Ward, A.C.; Azad, A.A.; Macreadie, I.G.; McKimm-Breeshkin, J.L.  
Nucleic Acids Res. 21, 2257, 1993  
A/Title: Complete nucleotide sequence of the non-structural gene of the human influenza A  
A/Reference number: S33685; MUID:93275765; PMID:8502573  
A/Accession: S33685  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: mRNA  
A/Residues: 1-121 <MAR>  
A/Cross-references: EMBL:Z21498; NID:9296585; PIDN:CAA79707.1; PID:9296586  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1993  
C/Genetics:  
A/Intons: 10/3  
C/Superfamily: influenza virus nonstructural protein NS2

Query Match  
5.1%; Score 7; DB 2; Length 121;  
Best Local Similarity 100.0%; Pred. No. 9.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLL 12  
| | | | |  
Db 101 QALQLL 107

RESULT 25  
A11058  
probable membrane protein STY4797 [imported] - Salmonella enterica subsp. enterica serov  
C/Species: Salmonella enterica subsp. enterica serovar Typh  
A/Note: this species has also been called Salmonella typh  
C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C/Accession: A11058  
R/Parikh, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
th, T.; Connor, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Garra, P.  
Nature 413, 848-852, 2001  
A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.  
A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A/Reference number: AB0502; MUID:21534947; PMID:11677608  
A/Accession: A11058  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-148 <PAR>  
A/Cross-references: GB:AL513382; PIDN:CAD06918.1; PID:G16505566; GSPDB:GN00176  
C/Genetics:  
A/Gene: STY4797

Query Match  
5.1%; Score 7; DB 2; Length 148;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 AILITLM 19  
| | | | |  
Db 71 AILITLM 77

RESULT 26  
B72717

hypothetical protein APE0277 - Aeropyrum pernix (strain K1)  
 C/Species: Aeropyrum pernix  
 C/Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
 C/Accession: B72717  
 R/Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Hatakeyama, Y.; Jin-no, K.; Takahara, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999  
 A/Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum  
 A/Reference number: A72450; MUID:99310339; PMID:10382966  
 A/Status: preliminary  
 A/Accession: B72717  
 A/Molecule type: DNA  
 A/Residues: 1-164 <RAW>  
 A/Cross-references: DDBJ:AP000059; NID:95103911; PIDN:BAA79230.1; PID:dl043016; PID:gs10  
 A/Experimental source: strain K1  
 C/Genetics:  
 A/Gene: APE0277

Query Match 5.1%; Score 7; DB 2; Length 164;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 LQLLAI 14  
 |||||  
 Db 157 LQLLAI 163

RESULT 27  
 D72061  
 hypothetical protein CP0175 [imported] - Chlamydia pneumoniae (strains CWL029 and AF  
 C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
 C/Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-May-2000  
 C/Accession: D72061; A81606  
 R/Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
 Nature Genet. 21, 385-389, 1999  
 A/Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
 A/Reference number: A72000; MUID:99206606; PMID:10192388  
 A/Status: preliminary  
 A/Accession: D72061  
 A/Molecule type: DNA  
 A/Residues: 1-184 <RAW>  
 A/Cross-references: GB:AE001642; GB:AE001363; NID:94376865; PIDN:AAD18714.1; PID:9437686  
 A/Experimental source: strain CWL029  
 R/Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
 C.; Dodson, R.; Gilm, M.; Nelson, W.; Deboy, J.; McClarty, G.; Salzberg,  
 Nucleic Acids Res. 28, 1397-1406, 2000  
 A/Title: Genome sequences of Chlamydia trachomatis Mohn and Chlamydia pneumoniae AR39.  
 A/Reference number: A81500; MUID:20150255; PMID:10684935  
 A/Status: preliminary  
 A/Accession: A81606  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-184 <REA>  
 A/Cross-references: GB:AE002178; GB:AE002161; NID:97189099; PIDN:AAF38049.1; PID:9718910  
 A/Experimental source: strain AR39, HL cells  
 C/Genetics:  
 A/Gene: CP0175

Query Match 5.1%; Score 7; DB 2; Length 184;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 RKTPLS 32  
 |||||  
 Db 26 RKTPLS 32

RESULT 28  
 B86562  
 hypothetical protein CP10574 [imported] - Chlamydia pneumoniae (strain J138)  
 C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae  
 C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
 C/Accession: B86562  
 R/Shitai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ito

Nucleic Acids Res. 28, 2311-2314, 2000  
 A/Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.  
 A/Reference number: A86491; MUID:2030349; PMID:10871362  
 A/Accession: B86562  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-184 <STO>  
 A/Cross-references: GB:BA000008; NID:98978945; PIDN:BAA98780.1; GSPDB:GN00142  
 A/Experimental source: strain J138  
 C/Genetics:  
 A/Gene: CP10574

Query Match 5.1%; Score 7; DB 2; Length 184;  
 Best Local Similarity 100.0%; Pred. No. 14;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 RKTPLS 32  
 |||||  
 Db 26 RKTPLS 32

RESULT 29  
 C71159  
 hypothetical protein PH0472 - Pyrococcus horikoshii  
 C/Species: Pyrococcus horikoshii  
 C/Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 28-Jul-2000  
 C/Accession: C71159  
 R/Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hatakeyama, Y.; Hino, Y.; Yamamoto, S.; Sekine,  
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,  
 DNA Res. 5, 55-76, 1998  
 A/Title: Complete sequence and gene organization of the genome of a hyper-thermophilic an  
 A/Reference number: A71000; MUID:98344137; PMID:9679194  
 A/Status: preliminary  
 A/Accession: C71159  
 A/Molecule type: DNA  
 A/Residues: 1-225 <RAW>  
 A/Cross-references: GB:AP000002; NID:93236129; PIDN:BAA29560.1; PID:93256877  
 A/Experimental source: strain OT3  
 A/Note: this accession replaces an interim accession for a sequence replaced by GenBank  
 C/Genetics:  
 A/Gene: PH0472

Query Match 5.1%; Score 7; DB 2; Length 225;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 70 LKVRQV 76  
 |||||  
 Db 10 LKVRQV 16

RESULT 30  
 C90915  
 hypothetical protein ECG2291 [imported] - Escherichia coli (strain O157:H7, substrain R1N  
 C/Species: Escherichia coli  
 C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
 C/Accession: C90915  
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
 gaewara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom  
 A/Reference number: A99629; MUID:21156231; PMID:11258796  
 A/Status: preliminary  
 A/Accession: C90915  
 A/Molecule type: DNA  
 A/Residues: 1-248 <RAY>  
 A/Cross-references: GB:BA000007; PIDN:BA035714.1; PID:913361757; GSPDB:GN00154  
 A/Experimental source: strain O157:H7, substrain R1MD 0509952  
 C/Genetics:  
 A/Gene: ECG2291  
 C/Superfamily: Escherichia coli ynfC protein

Query Match 5.1%; Score 7; DB 2; Length 248;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 LLAIIILT 17  
 |||||  
 Db 21 LLAIIILT 27

## RESULT 31

H85763

hypothetical protein ynfC [imported] - Escherichia coli (strain O157:H7, substrain EDL93  
 C/Species: Escherichia coli

C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C/Accession: H85763

R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grotbeck, E.O.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A/Reference number: A85480; MUID:21074935; PMID:11206551

A/Accession: H85763

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-248 &lt;STO&gt;

A/Cross-references: GB:AE005174; NID:g12515560; PIDN:AA656572.1; GSPDB:GN00145; UMG:225  
 A/Experimental source: strain O157:H7, substrain EDL933

C/Genetics:

A/Gene: ynfC

C/Superfamily: Escherichia coli ynfC protein

Query Match 5.1%; Score 7; DB 2; Length 248;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 LLAIIILT 17  
 |||||  
 Db 21 LLAIIILT 27

## RESULT 32

B83459

probable potaesium channel PA1496 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C/Species: Pseudomonas aeruginosa

C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C/Accession: B83459

R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
 .; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000

A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A/Reference number: A82950; MUID:20437337; PMID:10984043

A/Accession: B83459

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-283 &lt;STO&gt;

A/Cross-references: GB:AE004578; GB:AE004091; NID:g9947444; PIDN:AA604885.1; GSPDB:GN001  
 A/Experimental source: strain PA01

C/Genetics:

A/Gene: PA1496

Query Match 5.1%; Score 7; DB 2; Length 283;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 RIRFVLK 71  
 |||||  
 Db 123 RIRFVLK 129

## RESULT 33

C86687

hypothetical protein yefI [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
 C/Species: Lactococcus lactis subsp. lactis

C/Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
 C/Accession: C86687

R/Bolotin, A.; Winkler, P.; Mager, S.; Tallon, O.; Malarme, K.; Weisenbach, J.; Ehrlich  
 Genome Res. 11, 731-753, 2001

A/Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp

A/Reference number: A86625; MUID:21235186; PMID:11337471

A/Accession: C86687

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-288 &lt;STO&gt;

A/Cross-references: GB:AE005176; PID:g12723381; PIDN:AAK04597.1; GSPDB:GN00146  
 A/Experimental source: strain IL1403

C/Genetics:

A/Gene: yefI

Query Match 5.1%; Score 7; DB 2; Length 288;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 LLAIIILT 16  
 |||||  
 Db 21 LLAIIILT 27

## RESULT 34

S24386

protein kinase (BC 2.7.1.37) cd2 homolog - slime mold (Dictyostelium discoideum)

C/Species: Dictyostelium discoideum

C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 18-Jun-1999

C/Accession: S24386

R/Michaelis, C.; Weeks, G.

Biochim. Biophys. Acta 1132, 35-42, 1992

A/Title: Isolation and characterization of a cd2 cDNA from Dictyostelium discoideum.

A/Reference number: S24386; MUID:92379089; PMID:1311011

A/Accession: S24386

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-296 &lt;MTC&gt;

A/Cross-references: EMBL:M80808; NID:g167685; PIDN:AA33178.1; PID:g167686

C/Superfamily: kinase-related transforming protein; protein kinase homology

C/Keywords: ATP; phosphoprotein; phosphotransferase; serine/threonine-specific protein ki

F/8-257/Domain: protein kinase homology <KIN>

F/16-24/Region: protein kinase ATP-binding motif

F/39,55,129,131/Active site: Lys, Glu, Asp, Lys #status predicted

Query Match 5.1%; Score 7; DB 2; Length 296;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 64 FRIRFVL 70  
 |||||  
 Db 215 FRIRFVL 221

## RESULT 35

G83556

hypothetical protein PA0702 [imported] - Pseudomonas aeruginosa (strain PA01)

C/Species: Pseudomonas aeruginosa

C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C/Accession: G83556

R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
 .; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000

A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

A/Reference number: A82950; MUID:20437337; PMID:10984043

A/Accession: G83556

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-301 &lt;STO&gt;

A/Cross-references: GB:AE004506; GB:AE004091; NID:g9946584; PIDN:AA604091.1; GSPDB:GN001;  
 A/Experimental source: strain PA01

C/Genetics:

A:Gene: PA0702

Query Match

Best Local Similarity 100.0%; Pred. No. 21;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 LALLLT 14

Db 108 LALLLT 114

RESULT 36

P64905

probable sugar transport permease protein b1515 - Escherichia coli (strain K-12)

C/Species: Escherichia coli

C/Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002

C/Accession: F64905

R/Batner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A/Title: The complete genome sequence of Escherichia coli K-12.

A/Reference number: A64720; PMID:97426617; PMID:9278503

A/Accession: F64905

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-330 &lt;BLAT&gt;

A/Cross-references: GB:AE000249; GB:U00096; NID:G1787790; PIDN:AACT4588.1; PID:G1787794;

A/Experimental source: strain K-12, substrain MG1655

C/Superfamily: 1-arabinose transport system permease arah

C/Keywords: transmembrane protein

F/9-25/Domain: transmembrane #status predicted &lt;TM1&gt;

F/42-56/Domain: transmembrane #status predicted &lt;TM2&gt;

F/63-79/Domain: transmembrane #status predicted &lt;TM3&gt;

F/87-103/Domain: transmembrane #status predicted &lt;TM4&gt;

F/116-132/Domain: transmembrane #status predicted &lt;TM5&gt;

F/160-176/Domain: transmembrane #status predicted &lt;TM6&gt;

F/214-230/Domain: transmembrane #status predicted &lt;TM7&gt;

F/240-256/Domain: transmembrane #status predicted &lt;TM8&gt;

F/263-279/Domain: transmembrane #status predicted &lt;TM9&gt;

F/296-312/Domain: transmembrane #status predicted &lt;TM10&gt;

Query Match

Best Local Similarity 100.0%; Pred. No. 23;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 LALLLT 18

Db 88 LALLLT 94

RESULT 37

B90894

probable transport system permease protein Ecs2122 [imported] - Escherichia coli (strain

C/Species: Escherichia coli

C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001

C/Accession: B90894

R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A/Reference number: A9629; PMID:21156231; PMID:11258796

A/Accession: B90894

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-330 &lt;NAV&gt;

A/Cross-references: GB:BA000007; PIDN:BA835545.1; PID:G13361588; GSPDB:GN00154

C/Experimental source: strain O157:H7, substrain R1MD 0509552

C/Genetic:

A/Gene: Ecs2122

C/Superfamily: 1-arabinose transport system permease arah

Query Match

Best Local Similarity 100.0%; Pred. No. 23;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 LALLLT 18

Db 88 LALLLT 94

RESULT 38

G85723

probable transport system permease protein ydeZ [imported] - Escherichia coli (strain O1:

C/Species: Escherichia coli

C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C/Accession: G85723

R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,

Miller, L.; Grobeck, E.O.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A/Reference number: A85480; PMID:21074935; PMID:11206551

A/Accession: G85723

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-330 &lt;STO&gt;

A/Cross-references: GB:AE005174; NID:G12515151; PIDN:AA656251.1; GSPDB:GN00145; UWGP:Z215

C/Experimental source: strain O157:H7, substrain EDL933

C/Genetic:

A/Gene: ydeZ

C/Superfamily: 1-arabinose transport system permease arah

Query Match

Best Local Similarity 100.0%; Pred. No. 23;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 LALLLT 18

Db 88 LALLLT 94

RESULT 39

S41003

protein kinase (EC 2.7.1.37) cdc2 homolog - Caenorhabditis elegans

N/Alternate names: hypothetical protein T0565.3; p34; protein kinase MCC-1

C/Species: Caenorhabditis elegans

C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C/Accession: S41003; S26572; S52565

R/Thomas, K.

submitted to the EMBL Data Library, October 1993

A/Reference number: S41001

A/Accession: S41003

A/Molecule type: DNA

A/Residues: 1-332 &lt;THO&gt;

A/Cross-references: EMBL:Z27079; NID:G414641; PID:G414644

R/Ferraz, C.; Thierry-Mieg, D.; le Peuch, C.J.

submitted to the EMBL Data Library, September 1992

A/Description: Complete nucleotide sequence of a cDNA coding for a p34-cdc2-like protein

A/Reference number: S26572

A/Accession: S26572

A/Molecule type: mRNA

A/Residues: 1-332 &lt;FER&gt;

A/Cross-references: EMBL:X68384; NID:G6659; PIDN:CAA48455.1; PID:G6660

R/Mori, H.; Palmer, R.R.; Sternberg, P.W.

Mol. Gen. Genet. 245, 781-786, 1994

A/Title: The identification of a Caenorhabditis elegans homolog of p34(cdc2) kinase.

A/Reference number: S52565; PMID:95131956; PMID:7830726

A/Accession: S52565

A/Molecule type: mRNA

A/Status: preliminary

A/Cross-references: GB:S75262; NID:G807196; PIDN:AA60520.1; PID:G807197

A/Note: the sequence from Fig. 2 is inconsistent with that from Fig. 1 in having 135-Leu

A/Intons: 60/3; 129/3; 181/3; 290/3

C/Complex: In various organisms, cdc2 has been identified as a component of the M-phase

C/Superfamily: kinase-related transforming protein; protein kinase homolog

C/Keywords: ATP; cell cycle control; phosphoprotein; phosphotransferase; serine/threonine  
 F:20-27/Domain: protein kinase homology <KIN>  
 F:28-36/Region: protein kinase ATP-binding motif  
 F:51,69,146,148/Active site: Lys, Glu, Asp, Lys #status predicted

Query Match 5.1%; Score 7; DB 1; Length 332;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 232 FRIRVRL 238

RESULT 40  
 AG3026  
 hypothetical protein rbcC [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
 C/Species: Agrobacterium tumefaciens  
 C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
 C/Accession: AG3026

R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monk, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kucyavin, T.; Levy, R.; Li, M.; McClell  
 ; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
 ster, E.W.  
 A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A/Reference number: AB2577; MUID:21608550; PMID:11743193  
 A/Accession: AG3026  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-332 <KIR>  
 A/Cross-references: GB:AE008689; PIDN:ALA4629.1; PID:917742252; GSPDB:GN00187  
 A/Experimental source: strain C58 (Dupont)  
 C/Genetics:  
 A/Gene: rbcC  
 A/Map position: linear chromosome  
 C/Superfamily: 1-arabinose transport system permease araH

Query Match 5.1%; Score 7; DB 2; Length 332;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 10 LLAAILL 16  
 32 LLAAILL 38

RESULT 41  
 B98258  
 ribose ABC transporter, permease protein VCA0129 [imported] - Agrobacterium tumefaciens  
 C/Species: Agrobacterium tumefaciens  
 C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 18-Nov-2002  
 C/Accession: B98258  
 R;Gooder, B.; Hinkle, G.; Galtung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
 A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
 Science 294, 2323-2328, 2001  
 A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
 A/Reference number: A97359; MUID:21608551; PMID:11743194  
 A/Accession: B98258  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-332 <KUR>  
 A/Cross-references: GB:AE007870; PIDN:AKR9588.1; PID:915159477; GSPDB:GN00170  
 C/Genetics:  
 A/Gene: AGR L 2025  
 A/Map position: linear chromosome  
 C/Superfamily: 1-arabinose transport system permease araH

Query Match 5.1%; Score 7; DB 2; Length 332;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 LLAAILL 16  
 32 LLAAILL 38

RESULT 42  
 T20575  
 hypothetical protein F08B12.2 - Caenorhabditis elegans  
 C/Species: Caenorhabditis elegans  
 C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
 C/Accession: T20575

R;Dobson, R.  
 submitted to the EMBL Data Library, November 1995  
 A/Reference number: Z19295  
 A/Accession: T20575  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-359 <WIL>  
 A/Cross-references: EMBL:Z68104; PIDN:CAA92113.1; GSPDB:GN00028; CESP:F08B12.2  
 A/Experimental source: clone F08B12  
 C/Genetics:  
 A/Gene: CESP:F08B12.2  
 A/Map position: X  
 A/Intons: 47/3; 84/1; 148/3; 341/2

Query Match 5.1%; Score 7; DB 2; Length 359;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 28 KTFPSVH 34  
 55 KTFPSVH 61

RESULT 43  
 B71516  
 probable Fe-S oxidoreductase - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
 C/Species: Chlamydia trachomatis  
 C/Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999  
 C/Accession: B71516  
 R;Stephens, R.S.; Kalman, S.; Lammel, C.U.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,  
 Science 282, 754-759, 1998  
 A/Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tract  
 A/Reference number: A71570; MUID:99000809; PMID:9784136  
 A/Accession: B71516  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-369 <ARN>  
 A/Cross-references: GB:AE001316; GB:AE001273; NID:G3328850; PIDN:AA068023.1; PID:G3328850  
 A/Experimental source: serotype D, strain UW-3/Cx  
 C/Genetics:  
 A/Gene: CT426  
 C/Superfamily: hypothetical protein AF0390

Query Match 5.1%; Score 7; DB 2; Length 369;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 6 QATQLLL 12  
 32 QATQLLL 38

RESULT 44  
 F81674  
 conserved hypothetical protein TC0710 [imported] - Chlamydia muridarum (strain N199)  
 C/Species: Chlamydia muridarum, Chlamydia trachomatis MoKn  
 C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-May-2000  
 C/Accession: F81674  
 R;Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, J.  
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
 Nucleic Acids Res. 28, 1397-1406, 2000  
 A/Title: Genome sequences of Chlamydia trachomatis MoKn and Chlamydia pneumoniae AR39.

A:Reference number: AB1500; MUID:20150255; PMID:10684935  
A:Accession: F81674  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-369 <RT>  
A:Cross-references: GB:AB002339; GB:AB002160; NID:g7190728; PIDN:AAF39523.1; PID:g719074  
A:Experimental source: strain N19g (MOPN)  
A:Gene: TC0710  
C:Superfamily: hypothetical protein AF0390

Query Match 5.1%; Score 7; DB 2; Length 369;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 32 QALQLLL 12  
32 QALQLLL 38

RESULT 45  
F71234  
probable Na<sup>+</sup>/H<sup>+</sup>-exchanging protein - *Pyrococcus horikoshii*  
C:Species: *Pyrococcus horikoshii*  
C>Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 21-Jul-2000  
C:Accession: F71234  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
DNA Rep. 5, 55-76, 1998  
A>Title: Complete sequence and gene organization of the genome of a hyper-thermophilic  
A:Reference number: A71000; MUID:98344137; PMID:9679194  
A:Accession: F71234  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-375 <RAW>  
A:Cross-references: GB:AP000001; NID:g3236128; PIDN:BAA29205.1; PID:g3256522  
A:Experimental source: strain OT3  
A>Note: this accession replaces an interim accession for a sequence replaced by GenBank  
C:Gene: PH0136  
C:Superfamily: Aquifex aeolicus Na<sup>+</sup>/H<sup>+</sup>-exchanging protein nap1

Query Match 5.1%; Score 7; DB 2; Length 375;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIIIL 16  
Db 5 LLLAIIIL 11

RESULT 46  
T04914  
hypothetical protein T10114.160 - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 11-Jun-1999  
C:Accession: T04914  
R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X  
submitted to the Protein Sequence Database, April 1998  
A:Reference number: Z15589  
A:Accession: T04914  
A:Molecule type: DNA  
A:Residues: 1-386 <BBV>  
A:Cross-references: EMBL:AL021712  
A:Experimental source: cultivar Columbia; BAC clone T10114  
C:Gene: T10114.160

Query Match 5.1%; Score 7; DB 2; Length 386;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 RIRRVLK 71  
Db 15 RIRRVLK 21

RESULT 47  
S20035  
pulp protein - *Klebsiella pneumoniae*  
C:Species: *Klebsiella pneumoniae*  
C>Date: 19-Mar-1997 #sequence\_revision 18-Jul-1997 #text\_change 20-Aug-1999  
C:Accession: S20035; S77951  
R:Posset, O.; d'Enfert, C.; Reyse, I.; Pugsley, A.P.  
Mol. Microbiol. 6, 95-105, 1992  
A>Title: Pullulanase secretion in *Bacterichia coli* K-12 requires a cytoplasmic protein ar  
A:Reference number: S20034; MUID:92149318; PMID:1738317  
A:Accession: S20035  
A:Molecule type: DNA  
A:Residues: 1-381 <POS>  
A:Cross-references: EMBL:M32613  
R:Pugsley, A.P.; D'Enfert, C.; Reyse, I.; Wandersman, C.  
submitted to the EMBL Data Library, January 1991  
A:Reference number: S77951  
A:Accession: S77951  
A:Molecule type: DNA  
A:Residues: 1-140,161-401 <PUG>  
A:Cross-references: EMBL:M32613; NID:g149301; PIDN:AAA25128.1; PID:g149305  
C:Superfamily: secretion protein xcpS  
C:Keywords: transmembrane protein  
P:162-188/Domain: transmembrane #status predicted <TM1>  
P:218-234/Domain: transmembrane #status predicted <TM2>  
P:372-390/Domain: transmembrane #status predicted <TM3>

Query Match 5.1%; Score 7; DB 2; Length 401;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIIIL 16  
Db 223 LLLAIIIL 229

RESULT 48  
AC1163  
flagellar hook-associated protein 2 flid homolog lmo0707 [imported] - *Listeria monocytogenes*  
C:Species: *Listeria monocytogenes*  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AC1163  
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseurget, O.; Entian, K.D.; Fajth, H.;  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunz, F.; Kurapkai, G.; Madueno, E.; Maitournam, A.; Mat  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Poland, J.A.; Voss, H.; Wehlend,  
A>Title: Comparative genomics of *Listeria* species  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AC1163  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-429 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAC98785.1; PID:g16410096; GSPDB:GN00177  
A:Experimental source: strain BGD-e  
C:Gene: lmo0707

Query Match 5.1%; Score 7; DB 2; Length 429;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 ITDQYNK 56  
Db 396 ITDQYNK 402

## RESULT 49

AC1522  
 flagellar hook-associated protein 2 Flid homolog lin0715 [imported] - Listeria innocua  
 C:Species: Listeria innocua  
 C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
 C:Accession: AC1522  
 R:Gasser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baguero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fehli, H.; Science 294, 849-852, 2001  
 A:Authors: Krefit, J.; Kuhn, M.; Kunet, F.; Kurapkac, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schlueter, T.; Simoes, N.; Tietze, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A.; Title: Comparative genomics of Listeria species.  
 A:Reference number: AB1077; MUID:21537279; PMID:11679669  
 A:Accession: AC1522  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-429 <GLA>  
 A:Cross-references: GB:AL592022; PIDN:CAC95947.1; PID:G16413167; GSPDB:GN00178  
 A:Experimental source: strain Clp11262  
 C:Genetics:  
 A:Gene: lin0715

Query Match 5.1%; Score 7; DB 2; Length 429;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 ITDQYNK 56  
 |||||  
 Db 396 ITDQYNK 402

## RESULT 50

T46443  
 hypothetical protein DKFZp434F0427.1 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
 C:Accession: T46443  
 R:Bioecker, H.; Boecker, M.; Brandt, P.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S. submitted to the Protein Sequence Database, January 2000  
 A:Reference number: Z23032  
 A:Accession: T46443  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-435 <AAA>  
 A:Cross-references: EMBL:AL137446  
 A:Experimental source: adult testis; clone DKFZp434F0427  
 C:Genetics:  
 A:Note: DKFZp434F0427.1

Query Match 5.1%; Score 7; DB 2; Length 435;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ALQDLIA 13  
 |||||  
 Db 88 ALQDLIA 94

Search completed: January 21, 2004, 12:27:56  
 Job time : 23 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 12:20:21 ; Search time 17 Seconds  
(without alignments)  
378.980 Million cell updates/sec

Title: US-09-941-314-2

Perfect score: 137  
Sequence: 1 MAFPPQALQLLAHLITLMA.....VFAPVPEQYKINKSCSSD 137

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Database : Swissprot\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	137	100.0	137	1 CS11_HUMAN	Q9H112 homo sapien
2	7	5.1	118	1 VNS2_IACKJ	P08275 influenza a
3	7	5.1	118	1 VNS2_IADJ1	P08273 influenza a
4	7	5.1	118	1 VNS2_IADU3	P08271 influenza a
5	7	5.1	118	1 VNS2_IAMYN	P08269 influenza a
6	7	5.1	118	1 VNS2_IATRS	P08279 influenza a
7	7	5.1	121	1 VNS2_IANA	P30913 influenza a
8	7	5.1	121	1 VNS2_IACKB	P36350 influenza a
9	7	5.1	121	1 VNS2_IACKG	P30914 influenza a
10	7	5.1	121	1 VNS2_IACKQ	OS7269 influenza a
11	7	5.1	121	1 VNS2_IAD2	P03510 influenza a
12	7	5.1	121	1 VNS2_IADW	P03505 influenza a
13	7	5.1	121	1 VNS2_IADU2	OS7278 influenza a
14	7	5.1	121	1 VNS2_IAGUN	P33145 influenza a
15	7	5.1	121	1 VNS2_IAM6	P33145 influenza a
16	7	5.1	121	1 VNS2_IAM8	P33144 influenza a
17	7	5.1	121	1 VNS2_IAM9	P33147 influenza a
18	7	5.1	121	1 VNS2_IAMAO	P33148 influenza a
19	7	5.1	121	1 VNS2_IAP10	P33149 influenza a
20	7	5.1	121	1 VNS2_IAP11	P33150 influenza a
21	7	5.1	121	1 VNS2_IAP12	P33150 influenza a
22	7	5.1	121	1 VNS2_IATXB	P30915 influenza a
23	7	5.1	121	1 VNS2_IATXR	P08277 influenza a
24	7	5.1	121	1 KRE9_CANGA	Q74663 candida gla
25	7	5.1	296	1 CDC2_DICDI	P41112 dicystosell
26	7	5.1	330	1 YDE2_ECOLI	P77651 escherichia
27	7	5.1	332	1 CDC2_CAREL	P34556 caenorhabdi
28	7	5.1	359	1 PEXC_CAREL	Q91919 caenorhabdi
29	7	5.1	371	1 CYB_ANISC	Q48014 anilius scy
30	7	5.1	381	1 GSPF_KLEPN	P15745 klebsiella
31	7	5.1	415	1 TNR3_MOUSE	P50284 mus musculu
32	7	5.1	445	1 TBA1_SOYBN	P12459 glycine max
33	7	5.1	497	1 KCS1_MOUSE	O35173 mus musculu

34	7	5.1	497	1 KCS1_RAT	O88758 rattus norv
35	7	5.1	526	1 KCS1_HUMAN	O96KX3 homo sapien
36	7	5.1	650	1 MTRF_STRMU	O02425 streptococc
37	7	5.1	661	1 C180_HUMAN	O99467 homo sapien
38	7	5.1	1013	1 PRML_DROME	P82255 drosophila
39	7	5.1	3396	1 POLG_DENIS	P33478 d genome po
40	6	4.4	54	1 ATP8_PARLI	P12697 paracentrot
41	6	4.4	69	1 HUNE_APIME	P31504 apis mellif
42	6	4.4	77	1 EX78_CIOAB	O97hd1 clostridium
43	6	4.4	79	1 CYB_CORME	P16364 corcorax me
44	6	4.4	79	1 CYB_POMSU	P16362 pomatoscomu
45	6	4.4	93	1 DEPI_CAVPO	P11478 cavia porce
46	6	4.4	93	1 DEPI_CAVPO	P49112 cavia porce
47	6	4.4	93	1 FTSE_METJA	O58284 methanococc
48	6	4.4	103	1 FTSE_ECOLI	Q46894 escherichia
49	6	4.4	103	1 VMEW_SHVX	O04582 shalioi vir
50	6	4.4	113	1 YNFB_ECOLI	P76170 escherichia
51	6	4.4	114	1 PARA_TRYBB	P18764 trypanosoma
52	6	4.4	121	1 VNS2_IATAC	O02600 influenza a
53	6	4.4	121	1 VNS2_IATAC	P03503 influenza a
54	6	4.4	121	1 VNS2_IATAN	P21432 influenza a
55	6	4.4	121	1 VNS2_IACHI	P11619 influenza a
56	6	4.4	121	1 VNS2_IAPOM	P03506 influenza a
57	6	4.4	121	1 VNS2_IALB2	P26133 influenza a
58	6	4.4	121	1 VNS2_IATSS	P03504 influenza a
59	6	4.4	122	1 YP94_MYCPN	P75191 mycoplasma
60	6	4.4	123	1 YRDB_BACSU	O07080 bacillus su
61	6	4.4	123	1 CRCB_ECOL6	O8xb8 escherichia
62	6	4.4	127	1 CRCB_ECOLI	P37002 escherichia
63	6	4.4	127	1 CRCB_SALTY	O8xb9 salmonella
64	6	4.4	130	1 FLHE_ECOLI	P76297 escherichia
65	6	4.4	139	1 CS11_MOUSE	Q9d289 mus musculu
66	6	4.4	146	1 Y337_MYCPN	P75297 mycoplasma
67	6	4.4	146	1 CYTC_HUMAN	P01034 homo sapien
68	6	4.4	146	1 CYTC_MACMU	O19092 macaca mula
69	6	4.4	146	1 CYTC_SAIISC	O19093 saimiri sci
70	6	4.4	146	1 H2B_TORAC	P93354 nicotiana t
71	6	4.4	160	1 ATPD_SCHPO	Q9p66 schistosach
72	6	4.4	160	1 Y178_HABIN	P43961 haemophilus
73	6	4.4	162	1 YGFS_ECOLI	Q46819 escherichia
74	6	4.4	171	1 TATB_ECOL6	O8fb17 escherichia
75	6	4.4	171	1 TATB_ECOLI	O69415 escherichia

#### ALIGNMENTS

RESULT 1	CS11_HUMAN	STANDARD;	PRT;	137 AA.
ID	Q9H112; Q9H113;			
AC	Q9H112; Q9H113;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Cystatin II precursor.			
GN	CS11 OR CST1L.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).			
RX	MEDLINE=1638749; PubMed=11780052;			
RA	Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,			
RA	Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,			
RA	Batley J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,			
RA	Beasley O.P., Bird C.P., Blakey S.E., Bridgman A.M., Brown A.J.,			
RA	Buck D., Buttrill W.D., Butler A.P., Carder C., Carter N.P.,			
RA	Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,			
RA	Clegg S., Cobley V.E., Collier R.R., Connor R.E., Corby N.R.,			
RA	Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,			
RA	Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,			
RA	Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,			

RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Buckle E., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.R., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Leivasalio M.H., Levensha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McCormachle L.J., McElay K., McMurtry A.A.,  
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,  
 RA Skuse C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilmberg L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.,  
 RT "The DNA sequence and comparative analysis of human chromosome 20.",  
 RL Nature 414:865-871(2001).  
 CC -1- SUBCELLULAR LOCATION: Secreted (Potential).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q9H112-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q9H112-2; Sequence=VSP\_001260;  
 CC Note=No experimental confirmation available;  
 CC -1- SIMILARITY: BELONGS TO THE CYSTATIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; AL096677; CAC13170.1; -;  
 CC DR EMBL; AL096677; CAC17423.1; -;  
 CC DR HSSP; P01038; 1A90.  
 CC DR Genev; HGNC:15959; CST11.  
 CC DR InterPro; IPR00010; Cystatin.  
 CC DR Pfam; PF00031; cystatin; 1.  
 CC DR SMART; SMO0043; CY; 1.  
 CC DR PROSITE; PS00287; CYSTATIN; FALSE NEG.  
 CC KW Thiol protease inhibitor; Signal: Alternative splicing.  
 CC FT SIGNAL 1 25 POTENTIAL.  
 CC FT CHAIN 26 137 CYSTATIN 11.  
 CC FT SITE 75 79 SECONDARY AREA OF CONTACT (POTENTIAL).  
 CC FT DISULFID 93 101 BY SIMILARITY.  
 CC FT DISULFID 114 134 BY SIMILARITY.  
 CC FT CARBOHYD 131 131 N-LINKED (GLCNAc. .) (POTENTIAL).  
 CC FT VARSPLIC 76 110 Missing (in isoform 2).  
 CC FT /FTId=VSP\_001260.  
 CC SQ SEQUENCE 137 AA; 16375 MW; CS85C8C39A585C3B CRC64;  
 CC  
 CC Query Match 100.0%; Score 137; DB 1; Length 137;  
 CC Best Local Similarity 100.0%; Pred. No. 1,9e-133; Indels 0; Gaps 0;  
 CC Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 1 MAEPWQALQLLAILLTLMLPYQARKKTFISVHEVNAVENVAKDSLQMTDQYNKESDD 60  
 CC |  
 CC DB 1 MAEPWQALQLLAILLTLMLPYQARKKTFISVHEVNAVENVAKDSLQMTDQYNKESDD 60  
 CC |  
 CC QY 61 KYHFRIFVAVKQKQVTDHLEHVLNVEVQMTTCQKPEFTTNCVQPERELHAKVNCFFSIFA 120  
 CC |  
 CC DB 61 KYHFRIFVAVKQKQVTDHLEHVLNVEVQMTTCQKPEFTTNCVQPERELHAKVNCFFSIFA 120  
 CC |  
 CC QY 121 VPMFEQYKIINKSCSSD 137  
 CC |  
 CC DB 121 VPMFEQYKIINKSCSSD 137  
 CC |  
 CC RESULT 2  
 CC VNS2\_IACKJ

ID VNS2\_IACKJ STANDARD; PRT; 118 AA.  
 AC P08275;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Nonstructural protein NS2 (Fragment).  
 GN 8.  
 OS Influenza A virus (strain A/Chicken/Japan/24).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza A viruses; Influenzavirus A.  
 OK NCBI\_TaxId=11340;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87236215; PubMed=2954302;  
 RA Nakajima K., Nobusawa E., Ogawa T., Nakajima S.,  
 RT "Genetic divergence of the NS genes of avian influenza viruses.",  
 RL Virology 158:465-468(1987).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=NS2;  
 CC IsoId=P08275-1; Sequence=Displayed;  
 CC Name=NS1;  
 CC IsoId=P08274-1; Sequence=External;  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; M1561; AAA43505.1; -;  
 CC DR InterPro; IPR000968; Flu\_NS2.  
 CC DR Pfam; PF00601; Flu\_NS2; 1.  
 CC KW Nonstructural protein; Alternative splicing.  
 CC FT NON TER 1  
 CC SQ SEQUENCE 118 AA; 13925 MW; D6A8B4D34A163A46 CRC64;  
 CC  
 CC Query Match 5.1%; Score 7; DB 1; Length 118;  
 CC Best Local Similarity 100.0%; Pred. No. 5.9; Indels 0; Gaps 0;  
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 6 QALQLLL 12  
 CC |  
 CC DB 98 QALQLLL 104  
 CC |  
 CC RESULT 3  
 CC VNS2\_IADL1 STANDARD; PRT; 118 AA.  
 AC P08273;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Nonstructural protein NS2 (Fragment).  
 GN 8.  
 OS Influenza A virus (strain A/Duck/England/1/56).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza A viruses; Influenzavirus A.  
 OK NCBI\_TaxId=11354;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87236215; PubMed=2954302;  
 RA Nakajima K., Nobusawa E., Ogawa T., Nakajima S.,  
 RT "Genetic divergence of the NS genes of avian influenza viruses.",  
 RL Virology 158:465-468(1987).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=NS2;  
 CC IsoId=P08273-1; Sequence=Displayed;  
 CC Name=NS1;  
 CC IsoId=P08272-1; Sequence=External;

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16563; AAA43511.1; -
DR InterPro; IPR000968; Flu_NS2.
DR Pfam; PF00601; Flu_NS2.1.
DR Nonstructural protein; Alternative splicing.
PT NON_TER 1
SQ SEQUENCE 118 AA; 13990 MW; SCAAD9289DD70EC8 CRC64;

Query Match 5.1%; Score 7; DB 1; Length 118;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQALL 12
DB 98 QALQALL 104

RESULT 4
VNS2_IADU3 STANDARD; PRT; 118 AA.
ID VNS2_IADU3 STANDARD; PRT; 118 AA.
AC P08271;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nonstructural protein NS2 (Fragment).
GN 8.
OS Influenza A virus (strain A/Duck/Ukraine/1/63).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11374;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=67236215; PubMed=2954302;
RA Nakajima K., Nobusawa E., Ogawa T., Nakajima S.;
RT "Genetic divergence of the NS genes of avian influenza viruses.";
RL Virology 158:465-468(1987).
-1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=NS2;
IsoId=P08271-1; Sequence=Displayed;
Name=NS1;
IsoId=P08270-1; Sequence=External;
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16565; AAA43513.1; -
DR InterPro; IPR000968; Flu_NS2.
DR Pfam; PF00601; Flu_NS2.1.
DR Nonstructural protein; Alternative splicing.
PT NON_TER 1
SQ SEQUENCE 118 AA; 13992 MW; CIA2A9D2BF6556B CRC64;

Query Match 5.1%; Score 7; DB 1; Length 118;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQALL 12
DB 98 QALQALL 104
```

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M17070; AAA43549.1; -
DR InterPro; IPR000968; Flu_NS2.
DR Pfam; PF00601; Flu_NS2.1.
DR Nonstructural protein; Alternative splicing.
PT NON_TER 1
SQ SEQUENCE 118 AA; 14037 MW; C70BAD9AB371376 CRC64;

Query Match 5.1%; Score 7; DB 1; Length 118;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQALL 12
DB 98 QALQALL 104

RESULT 5
VNS2_IAMYN STANDARD; PRT; 118 AA.
ID VNS2_IAMYN STANDARD; PRT; 118 AA.
AC P08259;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nonstructural protein NS2 (Fragment).
GN 8.
OS Influenza A virus (strain A/Mynah/Haneda-Thai/76).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11444;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=67236215; PubMed=2954302;
RA Nakajima K., Nobusawa E., Ogawa T., Nakajima S.;
RT "Genetic divergence of the NS genes of avian influenza viruses.";
RL Virology 158:465-468(1987).
-1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=NS2;
IsoId=P08269-1; Sequence=Displayed;
Name=NS1;
IsoId=P08268-1; Sequence=External;
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M17070; AAA43549.1; -
DR InterPro; IPR000968; Flu_NS2.
DR Pfam; PF00601; Flu_NS2.1.
DR Nonstructural protein; Alternative splicing.
PT NON_TER 1
SQ SEQUENCE 118 AA; 14037 MW; C70BAD9AB371376 CRC64;

Query Match 5.1%; Score 7; DB 1; Length 118;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQALL 12
DB 98 QALQALL 104

RESULT 6
VNS2_IATRS STANDARD; PRT; 118 AA.
ID VNS2_IATRS STANDARD; PRT; 118 AA.
AC P08279;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nonstructural protein NS2 (Fragment).
GN NS OR 8.
OS Influenza A virus (strain A/Tern/South Africa/61).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11476;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=67236215; PubMed=2954302;
RA Nakajima K., Nobusawa E., Ogawa T., Nakajima S.;
RT "Genetic divergence of the NS genes of avian influenza viruses.";
RL Virology 158:465-468(1987).
-1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
```

```

CC      Name=NS2;
CC      Isoid=P08279-1; Sequence=Displayed;
CC      Name=NS1;
CC      Isoid=P08278-1; Sequence=External;
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; M16564; AAA43573.1; -;
CC      InterPro; IPR000968; Flu_NS2.
CC      Pfam; PF00601; Flu_NS2; 1.
CC      Nonstructural protein; Alternative splicing.
CC      NON TER 1
CC      SEQUENCE 118 AA; 13990 MW; C715D2501D3CF277 CRC64;

Query Match 5.1%; Score 7; DB 1; Length 118;
Best Local Similarity 100.0%; Pred. No. 5.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 QALQLL 12
    |||||
Db 98 QALQLL 104

RESULT 7
VNS2_IAANA STANDARD; PRT; 121 AA.
ID VNS2_IAANA
AC P30913;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nonstructural protein NS2.
GN 8.
OS Influenza A virus (strain A/Anas acuta/Primoje/695/76).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OC NCBI_TaxID=11339;
CC [1]
CC SEQUENCE FROM N.A.
CC MEDLINE=91306439; PubMed=1830182;
CC Ludwig S., Schultz U., Mandler J., Fitch W.M., Scholtissek C.;
CC "Phylogenetic relationship of the nonstructural (NS) genes of
CC Influenza A viruses.";
CC Virology 183:566-577(1991).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=NS2;
CC Isoid=P30913-1; Sequence=Displayed;
CC Name=NS1;
CC Isoid=P13138-1; Sequence=External;
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M60800; -; NOT ANNOTATED_CDS.
CC InterPro; IPR000256; Flu_NS1.
CC InterPro; IPR000968; Flu_NS2.
CC Pfam; PF00600; Flu_NS1; 1.
CC Pfam; PF00601; Flu_NS2; 1.
CC Nonstructural protein; Alternative splicing.
CC NON TER 1
CC SEQUENCE 121 AA; 14370 MW; 8FBAE09AE57362BB CRC64;

Query Match 5.1%; Score 7; DB 1; Length 121;

```

```

Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 QALQLL 12
    |||||
Db 101 QALQLL 107

RESULT 8
VNS2_IACKB STANDARD; PRT; 121 AA.
ID VNS2_IACKB
AC P36350;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nonstructural protein NS2.
GN 8.
OS Influenza A virus (strain A/Chicken/Brescia/1902).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OC NCBI_TaxID=36418;
CC [1]
CC SEQUENCE FROM N.A.
CC MEDLINE=92109567; PubMed=1530908;
CC Klimov A., Proesch S., Schaefer J., Bucher D.;
CC "Subtype H7 influenza viruses: comparative antigenic and molecular
CC analysis of the HA-, M-, and NS-genes.";
CC Arch. Virol. 122:143-161(1992).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=NS2;
CC Isoid=P36350-1; Sequence=Displayed;
CC Name=NS1;
CC Isoid=P36349-1; Sequence=External;
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; I37795; -; NOT ANNOTATED_CDS.
CC InterPro; IPR000256; Flu_NS1.
CC InterPro; IPR000968; Flu_NS2.
CC Pfam; PF00600; Flu_NS1; 1.
CC Pfam; PF00601; Flu_NS2; 1.
CC Nonstructural protein; Alternative splicing.
CC NON TER 1
CC SEQUENCE 121 AA; 14334 MW; FB6B839AE070E13A CRC64;

Query Match 5.1%; Score 7; DB 1; Length 121;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 QALQLL 12
    |||||
Db 101 QALQLL 107

RESULT 9
VNS2_IACKG STANDARD; PRT; 121 AA.
ID VNS2_IACKG
AC P30914;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nonstructural protein NS2.
GN 8.
OS Influenza A virus (strain A/Chicken/Germany/n/49).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OC NCBI_TaxID=11339;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=91306439; PubMed=1830182;
RW Ludwig S., Schultz U., Mandler J., Fitch W.M., Scholtissek C.;
RT "Phylogenetic relationship of the nonstructural (NS) genes of
RT Influenza A viruses."
RL Virology 183:566-577(1991).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named Isoforms=2;
CC Name=NS2;
CC IsoId=P30914-1; Sequence=Displayed;
CC Name=NS1;
CC IsoId=P30909-1; Sequence=External;
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M55464; -; NOT_ANNOTATED_CDS.
DR InterPro; IPR000256; Flu_NS1.
DR InterPro; IPR000968; Flu_NS2.
DR Pfam; PF00600; Flu_NS1; 1.
DR Pfam; PF00601; Flu_NS2; 1.
DR Nonstructural protein; Alternative splicing.
SQ SEQUENCE 121 AA; 14314 MW; C0F5CE50CA87222F4 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLL 12
DB 101 QALQLL 107

RESULT 10
VNS2_IACKO STANDARD; PRT; 121 AA.
AC 057269;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nonstructural protein NS2.
GN NS OR 8.
OS Influenza A virus (strain A/Chicken/Pennsylvania/1370/83).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OK NCBI_TaxID=11342;
RN (1)
RP SEQUENCE FROM N.A.
RA Suarez D.L.;
RT "Comparison of avian influenza nonstructural gene sequences."
RT Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named Isoforms=2;
CC Name=NS2;
CC IsoId=O57269-1; Sequence=Displayed;
CC Name=NS1;
CC IsoId=O57268-1; Sequence=External;
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U96739; AAB93937.1; -

```

```

DR InterPro; IPR000256; Flu_NS1.
DR InterPro; IPR000968; Flu_NS2.
DR Pfam; PF00600; Flu_NS1; 1.
DR Pfam; PF00601; Flu_NS2; 1.
RW Nonstructural protein; Alternative splicing.
SQ SEQUENCE 121 AA; 14305 MW; B560752156684892 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLL 12
DB 101 QALQLL 107

RESULT 11
VNS2_IADR2 STANDARD; PRT; 121 AA.
AC P03510; P13151;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Nonstructural protein NS2.
GN 8.
OS Influenza A virus (strain A/Duck/Alberta/60/76),
OS Influenza A virus (strain A/Pittail/Alberta/358/79), and
OS Influenza A virus (strain A/Turkey/Canada/63).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OK NCBI_TaxID=11347, 11452, 31664;
RN (1)
RP SEQUENCE FROM N.A.
RA STRAIN=A/Duck/Alberta/60/76;
RC MEDLINE=81276929; PubMed=6927848;
RA Baez M., Zazra J.J., Elliott R.M., Young J.F., Palase P.;
RT "Nucleotide sequence of the influenza A/duck/Alberta/60/76 virus NS
RT RNA: conservation of the NS1/NS2 overlapping gene structure in a
RT divergent influenza virus RNA segment."
RL Virology 113:397-402(1981).
RN (2)
RP SEQUENCE FROM N.A.
RA STRAIN=A/Pittail/Alberta/358/79;
RC MEDLINE=89299445; PubMed=252836;
RA Treanor J.J., Snyder M.H., London W.T., Murphy B.R.;
RT "The B allele of the NS gene of avian influenza viruses, but not the
RT A allele, attenuates a human influenza A virus for squirrel
RT monkeys."
RL Virology 171:1-9(1989).
RN (3)
RP SEQUENCE FROM N.A.
RA STRAIN=A/Turkey/Canada/63;
RC MEDLINE=91306439; PubMed=1830182;
RW Ludwig S., Schultz U., Mandler J., Fitch W.M., Scholtissek C.;
RT "Phylogenetic relationship of the nonstructural (NS) genes of
RT Influenza A viruses."
RT Virology 183:566-577(1991).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named Isoforms=2;
CC Name=NS2;
CC IsoId=P03510-1; Sequence=Displayed;
CC Name=NS1;
CC IsoId=P13143-1; Sequence=External;
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J02105; AAA43508.1; -

```

DR EMBL; M25370; AAA43558.1; -.  
 DR EMBL; M55468; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; A04098; MNIIV26.  
 DR InterPro; IPR000256; Flu\_NS1.  
 DR InterPro; IPR000968; Flu\_NS2.  
 DR Pfam; PF00600; Flu\_NS1; 1.  
 DR Pfam; PF00601; Flu\_NS2; 1.  
 DR Nonstructural protein; Alternative splicing.  
 SQ SEQUENCE 121 AA; 14287 MW; D1143B50CA9A3FF4 CRC64;

Query Match 5.1%; Score 7; DB 1; Length 121;  
 Best Local Similarity 100.0%; Pred. No. 6;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 QALQALL 12  
 |||||  
 Db 101 QALQALL 107

RESULT 12  
 VNS2\_IAGUN STANDARD; PRT; 121 AA.  
 AC P03505; Sequence=Not described;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Nonstructural protein NS2.  
 GN 8.  
 OS Influenza A virus (strain A/Port Warren/1/50).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza A viruses; Influenzavirus A.  
 NCBI\_TaxID=11381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 83164298; PubMed=6834468;  
 RA Kysaral M., Buonagurio D.A., Young J.F., Palese P.;  
 RT "Sequential mutations in the NS genes of Influenza virus field  
 strains";  
 RL J. Virol. 45:547-554(1983).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=NS2;  
 CC IsoId=P03505-1; Sequence=Displayed;  
 CC Name=NS1;  
 CC IsoId=P03497-1; Sequence=External;  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U02087; AAA43521.1; -.  
 DR EMBL; U02087; AAA43521.1; -.  
 DR InterPro; IPR000256; Flu\_NS1.  
 DR InterPro; IPR000968; Flu\_NS2.  
 DR Pfam; PF00600; Flu\_NS1; 1.  
 DR Pfam; PF00601; Flu\_NS2; 1.  
 DR Nonstructural protein; Alternative splicing.  
 SQ SEQUENCE 121 AA; 14433 MW; 02D85B6D83BAE14 CRC64;

Query Match 5.1%; Score 7; DB 1; Length 121;  
 Best Local Similarity 100.0%; Pred. No. 6;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 QALQALL 12  
 |||||  
 Db 101 QALQALL 107

RESULT 13

VNS2\_IAGU2 STANDARD; PRT; 121 AA.  
 ID VNS2\_IAGU2  
 AC 057278;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Nonstructural protein NS2.  
 GN NS OR 8.  
 OS Influenza A virus (strain A/Maryland/704/77), and  
 OS Influenza A virus (strain A/Gull/Massachusetts/26/80).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza A viruses; Influenzavirus A.  
 NCBI\_TaxID=11387, 11391;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Suarez D.L.;  
 RT "Comparison of avian influenza nonstructural gene sequences";  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=NS2;  
 CC IsoId=057278-1; Sequence=Displayed;  
 CC Name=NS1;  
 CC IsoId=057278-2; Sequence=Not described;  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U96737; AAB93934.1; -.  
 DR EMBL; U96744; AAB93948.1; -.  
 DR InterPro; IPR000256; Flu\_NS1.  
 DR InterPro; IPR000968; Flu\_NS2.  
 DR Pfam; PF00600; Flu\_NS1; 1.  
 DR Pfam; PF00601; Flu\_NS2; 1.  
 DR Nonstructural protein; Alternative splicing.  
 SQ SEQUENCE 121 AA; 14291 MW; 20B8B8462657640 CRC64;

Query Match 5.1%; Score 7; DB 1; Length 121;  
 Best Local Similarity 100.0%; Pred. No. 6;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 QALQALL 12  
 |||||  
 Db 101 QALQALL 107

RESULT 14  
 VNS2\_IAGUN STANDARD; PRT; 121 AA.  
 ID VNS2\_IAGUN  
 AC 057266;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Nonstructural protein NS2.  
 GN NS OR 8.  
 OS Influenza A virus (strain A/Gull/Minnesota/945/80).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza A viruses; Influenzavirus A.  
 NCBI\_TaxID=11392;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Suarez D.L.;  
 RT "Comparison of avian influenza nonstructural gene sequences";  
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=NS2;  
 CC IsoId=057266-1; Sequence=Displayed;

```

CC      Name=NS1;
CC      IsoId=O57266-2; Sequence=Not described;
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: U06738; AAB93936.1; -
CC      InterPro: IPR000968; Flu_NS2.
CC      Pfam: PF00601; Flu_NS2; 1.
CC      Nonstructural protein; Alternative splicing.
CC      SEQUENCE 121 AA; 14308 MW; 9405E5B0CA22B6F CRC64;
CC
CC      Query Match          5.1%; Score 7; DB 1; Length 121;
CC      Best Local Similarity 100.0%; Pred. No. 6;
CC      Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
Qy      6 QALQLL 12
Db      101 QALQLL 107

RESULT 15
VNS2_IAM6      STANDARD;      PRT;      121 AA.
ID_VNS2_IAM6
AC_P1315;
DT_01-JAN-1990 (Rel. 13, Created)
DT_01-MAY-1991 (Rel. 18, Last sequence update)
DT_28-FEB-2003 (Rel. 41, Last annotation update)
DE_Nonstructural protein NS2.
GN_8
OS_Influenza A virus (strain A/Mallard/Alberta/88/76).
OC_Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC_Influenza A viruses; Influenzavirus A.
OX_NCBI_TaxId=11430;
[1]
RP_SEQUENCE FROM N.A.
RX_MEDLINE=89299445; PubMed=2525836;
RA_Treanor J.J., Snyder M.H., London W.T., Murphy B.R.;
RT_The B allele of the NS gene of avian influenza viruses, but not the
RT_A allele, attenuates a human influenza A virus for squirrel
RT_monkeys.";
RT_Virology 171:1-9(1989).
RL_1- ALTERNATIVE PRODUCTS;
CC_Event=Alternative splicing; Named isoforms=2;
CC_Name=NS2;
CC_IsoId=P13145-1; Sequence=Displayed;
CC_Name=NS1;
CC_IsoId=P13137-1; Sequence=External;
CC-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: M25373; AAA43530.1; -
CC      InterPro: IPR000256; Flu_NS1.
CC      InterPro: IPR000968; Flu_NS2.
CC      Pfam: PF00600; Flu_NS1; 1.
CC      Pfam: PF00601; Flu_NS2; 1.
CC      Nonstructural protein; Alternative splicing.
CC      SEQUENCE 121 AA; 14273 MW; 210B3E58D6FCE499 CRC64;
CC
CC      Query Match          5.1%; Score 7; DB 1; Length 121;
CC      Best Local Similarity 100.0%; Pred. No. 6;
CC      Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      6 QALQLL 12
Db      101 QALQLL 107

RESULT 16
VNS2_IAM8      STANDARD;      PRT;      121 AA.
ID_VNS2_IAM8
AC_P1314;
DT_01-JAN-1990 (Rel. 13, Created)
DT_01-MAY-1991 (Rel. 18, Last sequence update)
DT_15-SEP-2003 (Rel. 42, Last annotation update)
DE_Nonstructural protein NS2.
GN_8
OS_Influenza A virus (strain A/Mallard/Alberta/82/78).
OC_Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC_Influenza A viruses; Influenzavirus A.
OX_NCBI_TaxId=11432;
[1]
RP_SEQUENCE FROM N.A.
RX_MEDLINE=89299445; PubMed=2525836;
RA_Treanor J.J., Snyder M.H., London W.T., Murphy B.R.;
RT_The B allele of the NS gene of avian influenza viruses, but not the
RT_A allele, attenuates a human influenza A virus for squirrel
RT_monkeys.";
RT_Virology 171:1-9(1989).
RL_1- ALTERNATIVE PRODUCTS;
CC_Event=Alternative splicing; Named isoforms=2;
CC_Name=NS2;
CC_IsoId=P13144-1; Sequence=Displayed;
CC_Name=NS1;
CC_IsoId=P03501-1; Sequence=External;
CC-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: M25372; AAA43528.1; -
CC      PIR: D32663; KNIVB6.
CC      PDB: 1LCQ; 17-APR-02.
CC      InterPro: IPR000256; Flu_NS1.
CC      InterPro: IPR000968; Flu_NS2.
CC      Pfam: PF00600; Flu_NS1; 1.
CC      Pfam: PF00601; Flu_NS2; 1.
CC      Nonstructural protein; Alternative splicing; 3D-structure.
CC      SEQUENCE 121 AA; 14315 MW; D1170E63CA9A3FP4 CRC64;
CC
CC      Query Match          5.1%; Score 7; DB 1; Length 121;
CC      Best Local Similarity 100.0%; Pred. No. 6;
CC      Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
Qy      6 QALQLL 12
Db      101 QALQLL 107

RESULT 17
VNS2_IAMN      STANDARD;      PRT;      121 AA.
ID_VNS2_IAMN
AC_P2157;
DT_01-MAY-1991 (Rel. 18, Created)
DT_01-MAY-1991 (Rel. 18, Last sequence update)
DT_28-FEB-2003 (Rel. 41, Last annotation update)
DE_Nonstructural protein NS2.
GN_8
OS_Influenza A virus (strain A/Mallard/New York/6750/78).
OC_Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC_Influenza A viruses; Influenzavirus A.

```

```

OK NCBI_TaxID=11435;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89299445; PubMed=2525836;
RA Treanor J.J., Snyder W.H., London W.T., Murphy B.R.;
RT "The B allele of the NS gene of avian influenza viruses, but not the
RT A allele, attenuates a human influenza A virus for squirrel
RT monkeys.";
RL Virology 171:1-9(1989).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=NS2;
CC IsoId=P131527-1; Sequence=Displayed;
CC Name=NS1;
CC IsoId=P13138-1; Sequence=External;
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ldb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL; M25376; AAA43546.1; -
DR InterPro; IPR000256; Flu_NS1.
DR InterPro; IPR000968; Flu_NS2.
DR Pfam; PF00600; Flu_NS1; 1.
DR Pfam; PF00601; Flu_NS2; 1.
KW Nonstructural protein; Alternative splicing.
SQ SEQUENCE 121 AA; 14338 MW; F97DE09AE57377AF CRC64;

Query Match 5.1%; Score 7; DB 1; Length 121;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLL 12
DB 101 QALQLL 107

RESULT 18
VNS2_IAMAO STANDARD; PRT; 121 AA.
AC P13147;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nonstructural protein NS2.
GN 8.
OS Influenza A virus (strain A/Mallard/New York/6874/79).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OC NCBI_TaxID=11436;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89299445; PubMed=2525836;
RA Treanor J.J., Snyder W.H., London W.T., Murphy B.R.;
RT "The B allele of the NS gene of avian influenza viruses, but not the
RT A allele, attenuates a human influenza A virus for squirrel
RT monkeys.";
RL Virology 171:1-9(1989).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=NS2;
CC Name=NS1;
CC IsoId=P13147-1; Sequence=Displayed;
CC Name=NS1;
CC IsoId=P13139-1; Sequence=External;
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ldb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ldb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL; M25375; AAA43544.1; -
DR InterPro; IPR000256; Flu_NS1.
DR InterPro; IPR000968; Flu_NS2.
DR Pfam; PF00600; Flu_NS1; 1.
DR Pfam; PF00601; Flu_NS2; 1.
KW Nonstructural protein; Alternative splicing.
SQ SEQUENCE 121 AA; 14306 MW; E4D7579AE56733FF CRC64;

Query Match 5.1%; Score 7; DB 1; Length 121;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLL 12
DB 101 QALQLL 107

RESULT 19
VNS2_IAP10 STANDARD; PRT; 121 AA.
AC P13148;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nonstructural protein NS2.
GN NS OR 8.
OS Influenza A virus (strain A/Piscataway/119/79).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OC NCBI_TaxID=11449;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89299445; PubMed=2525836;
RA Treanor J.J., Snyder W.H., London W.T., Murphy B.R.;
RT "The B allele of the NS gene of avian influenza viruses, but not the
RT A allele, attenuates a human influenza A virus for squirrel
RT monkeys.";
RL Virology 171:1-9(1989).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=NS2;
CC Name=NS1;
CC IsoId=P13148-1; Sequence=Displayed;
CC Name=NS1;
CC IsoId=P13140-1; Sequence=External;
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ldb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL; M25374; AAA43556.1; -
DR InterPro; IPR000256; Flu_NS1.
DR InterPro; IPR000968; Flu_NS2.
DR Pfam; PF00600; Flu_NS1; 1.
DR Pfam; PF00601; Flu_NS2; 1.
KW Nonstructural protein; Alternative splicing.
SQ SEQUENCE 121 AA; 14347 MW; ED8D049AE568AC6F CRC64;

Query Match 5.1%; Score 7; DB 1; Length 121;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLL 12
DB 101 QALQLL 107

```

```

RESULT 20
VNS2_IAP11 STANDARD; PRT; 121 AA.
ID VNS2_IAP11
AC P13149;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Nonstructural protein NS2.
GN NS OR 8.
OS Influenza A virus (strain A/Pintail/Alberta/121/79), and
OC Influenza A virus (strain A/Tern/Turkmenia/18/72).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
NCBI_TaxID=11450, 11477;
(1)
SEQUENCE FROM N.A.
RP STRAIN=A/Pintail/Alberta/121/79;
RX MEDLINE=89299445; PubMed=2525836;
RA Treanor J.J., Snyder M.H., London W.T., Murphy B.R.;
RT "The B allele of the NS gene of avian influenza A virus for squirrel
RT monkeys.";
RL Virology 171:1-9(1989).
(2)
SEQUENCE FROM N.A.
RP STRAIN=A/Tern/Turkmenia/18/72;
RX MEDLINE=91306439; PubMed=1830182;
RA Ludwig S., Schultz U., Mandler J., Fitch W.M., Scholtissek C.;
RT "Phylogenetic relationship of the nonstructural (NS) genes of
RT influenza A viruses.";
RL Virology 183:566-577(1991).
-1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=NS2;
IsoId=P13149-1; Sequence=Displayed;
Name=NS1;
IsoId=P13141-1, P30912-1; Sequence=External;
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; M25371; AAA43560.1; -
DR EMBL; M55466; -; NOT_ANNOTATED_CDS.
DR InterPro; IPR000256; Flu_NS1.
DR InterPro; IPR000968; Flu_NS2.
DR Pfam; PF00600; Flu_NS1; 1.
DR Pfam; PF00601; Flu_NS2; 1.
KW Nonstructural protein; Alternative splicing.
SQ SEQUENCE 121 AA; 14272 MW; 3BFA2FBBC4793324 CRC64;

Query Match 5.1%; Score 7; DB 1; Length 121;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLLL 12
DB 101 QALQLLL 107

```

```

RESULT 22
VNS2_IATK3 STANDARD; PRT; 121 AA.
ID VNS2_IATK3
AC P30915;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nonstructural protein NS2.
GN NS OR 8.
OS Influenza A virus (strain A/Turkey/Bechelem-G1111c/1492-B/82).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
NCBI_TaxID=11663;
(1)
SEQUENCE FROM N.A.
RP MEDLINE=91306439; PubMed=1830182;
RA Ludwig S., Schultz U., Mandler J., Fitch W.M., Scholtissek C.;
RT "Phylogenetic relationship of the nonstructural (NS) genes of
RT influenza A viruses.";
RL Virology 183:566-577(1991).
-1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=NS2;
IsoId=P30915-1; Sequence=Displayed;
Name=NS1;
IsoId=P30910-1; Sequence=External;
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation on its
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; M55467; -; NOT_ANNOTATED_CDS.
DR InterPro; IPR0000256; Flu_NS1.
DR InterPro; IPR000968; Flu_NS2.
DR Pfam; PF00600; Flu_NS1; 1.
DR Pfam; PF00601; Flu_NS2; 1.
DR Nonstructural protein; Alternative splicing.
KW SEQUENCE 121 AA; 14300 MW; 28B84B8EC4793324 CRC64;

Query Match 5.1%; Score 7; DB 1; Length 121;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQALL 12
101 QALQALL 107

RESULT 23
VNS2_IATKR STANDARD; PRT; 121 AA.
ID_VNS2_IATKR
AC P08277;
DT 01-AUG-1998 (Rel. 08, Created)
DT 01-AUG-1998 (Rel. 08, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nonstructural protein NS2.
NS OR 8.
GN Influenza A virus (strain A/Turkey/Oregon/71).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_TaxID=11472;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8712162; PubMed=3811235;
RX Norton G.P., Tanaka T., Tobita K., Nakada S., Buonagurio D.A.,
RX Greenspan D., Krystal M., Palese P.;
RA "Infectious influenza A and B virus variants with long carboxyl
RT terminal deletions in the NS1 polypeptides.";
RL Virology 156:204-213 (1987).
-1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=NS2;
IsoId=P08277-1; Sequence=Displayed;
Name=NS1;
IsoId=P08276-1; Sequence=External;
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; M16623; AAA43550.1; -
DR InterPro; IPR0000256; Flu_NS1.
DR InterPro; IPR000968; Flu_NS2.
DR Pfam; PF00600; Flu_NS1; 1.
DR Pfam; PF00601; Flu_NS2; 1.
DR Nonstructural protein; Alternative splicing.
KW SEQUENCE 121 AA; 14253 MW; 8B106F15DA9F62CC CRC64;

Query Match 5.1%; Score 7; DB 1; Length 121;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQALL 12

```

```

Db          101 QALQELL 107
|||||
|||||

RESULT 24
KRE9_CANGA STANDARD; PRT; 276 AA.
ID_KRE9_CANGA
AC 074683;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Cell wall synthesis protein KRE9 precursor.
GN KRE9.
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota, Fungi, Ascomycota, Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OC NCBI_TaxID=5478;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 2001;
RX MEDLINE=98422452; PubMed=9748432.
RA Nagashima S., Lusser W., Bussey H.;
RT "Isolation of Candida glabrata homologs of the Saccharomyces
RT cerevisiae KRE9 and KNL1 genes and their involvement in cell wall
RT beta-1,6-glucan synthesis."
RL J. Bacteriol. 180:5020-5029(1998).
CC -1- FUNCTION: INVOLVED IN CELL WALL BETA(1->6) GLUCAN SYNTHESIS.
CC -1- SUBCELLULAR LOCATION: SECRETED, PROBABLY FOUND AT CELL SURFACE (BY
CC SIMILARITY).
CC -1- PTM: O-GLYCOSYLATED (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE KRE9 / KNL1 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ch/announce/
CC or send an email to license@ebi-sib.ch).
-----
CC EMBL; AF064251; AAC64008.1; -.
DR KRE9_CANGA
KW GLYCOPROTEIN; Cell wall; Signal.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 276 CELL WALL SYNTHESIS PROTEIN KRE9.
SQ SEQUENCE 276 AA; 30549 MW; 5AA4328C8DC59EDD CRC64;
-----
Query Match 5.1%; Score 7; DB 1; Length 276;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLAAILL 16
|||||
|||||
Db 2 LLAAILL 8

RESULT 25
CDC2_DICDI STANDARD; PRT; 236 AA.
ID_CDC2_DICDI
AC P34112;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cell division control protein 2 homolog (EC 2.7.1.-) (p34 protein
DE kinase).
GN CDCB OR DD 01782.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota, Mycetozoa; Dictyostelida; Dictyostelium.
OC NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92379089; PubMed=1511011;
RA Michaelis C., Weeks G.;
RT "Isolation and characterization of a cdc 2 cDNA from Dictyostelium
```

```

RT discoidum."
RL Biochim. Biophys. Acta 1132:35-42(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafrański K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Thunberg B., Cox B., Quail M.A., Platzer M., Rosenthal A.,
RA Noegel A.A.;
RT "Sequence and analysis of chromosome 2 of Dictyostelium."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PLAYS A KEY ROLE IN THE CONTROL OF THE EUKARYOTIC
CC CELL CYCLE. IT IS REQUIRED IN HIGHER CELLS FOR ENTRY INTO S-PHASE
CC AND MITOSIS. P34 IS A COMPONENT OF THE KINASE COMPLEX THAT
CC PHOSPHORYLATES THE REPETITIVE CARBOXYL-TERMINUS OF RNA
CC POLYMERASE II.
CC -1- ENZYME REGULATION: PHOSPHORYLATION AT THR-20 OR TYR-21 INACTIVATES
CC THE ENZYME, WHILE PHOSPHORYLATION AT THR-162 ACTIVATES IT (BY
CC SIMILARITY).
CC -1- SUBUNIT: FORMS A STABLE BUT NON-COVALENT COMPLEX WITH A REGULATORY
CC SUBUNIT AND WITH A CYCLIN.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC CDC2/CDCX SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; MG0808; AAA3178.1; -.
DR EMBL; AC117076; AAM33721.1; -.
DR PIR; S24386; S24386.
DR HSSP; 000534; 1B18.
DR DictyDB; DD05040; cdcB.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR Pfam; PF00069; pkinase.1.
DR ProDom; PD000001; Prot_kinase.1.
DR SMART; SM00920; S_TKc_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP_1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST_1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM_1.
DR Transferrase; Serine/threonine-protein kinase; ATP-binding;
KM Cell cycle; Cell division; Mitosis; Phosphorylation.
KW DOMAIN
FT 10 288 PROTEIN KINASE.
FT NE_BIND 16 24 ATP (BY SIMILARITY).
FT BINDING 39 39 BY SIMILARITY.
FT ACT_SITE 129 129 BY SIMILARITY.
FT MOD_RES 20 20 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 21 21 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 162 162 PHOSPHORYLATION (BY CAY) (BY SIMILARITY).
SQ SEQUENCE 296 AA; 33773 MW; 406B53820AA677B6 CRC64;
Query Match 5.1%; Score 7; DB 1; Length 296;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 64 FRIFRVL 70
DB 215 FRIFRVL 221
RESULT 26
YDEZ_ECOLI
ID YDEZ_ECOLI STANDARD; PRT; 330 AA.
AC P77651; P76881;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical ABC transporter permease protein ydeZ.

```

```

GN YDEZ OR B1515 OR SF1584.1.
OS Escherichia coli, and
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 623;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=X12 / MG1655;
RX MEDLINE=9742617; PubMed=9278503;
RA Blattner P.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1234-1238(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=X12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Bada T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sampei G., Seki Y., Sivasubram S.,
RA Tengan H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horikuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map."
RL DNA Res. 3:363-377(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157."
RL Nucleic Acids Res. 30:4432-4441(2002).
CC -1- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM EGO/YDEZ. PROBABLY RESPONSIBLE FOR THE TRANSLOCATION OF
CC THE SUBSTRATE ACROSS THE MEMBRANE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM PERMEASE FAMILY. AAAH/RBSC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000249; AAC74588.1; -.
DR EMBL; D90794; BAA15203.1; -.
DR EMBL; D90795; BAA15203.1; -.
DR EMBL; AE015180; AAN43172.1; -.
DR PIR; F64905; F64905.
DR EcoGene; EG33808; ydeZ.
DR InterPro; IPR001851; Bac_innem_transp.
DR InterPro; IPR000515; BPD_innem.
DR Pfam; PF02653; BPD_transp_2; 1.
DR PROSITE; PS00402; BPD_TRANS_P. INN_MEMBER. FALSE NEG.
KM Hypothetical protein; Transport; Transmembrane; Inner membrane;
FT TRANSMEM 5 25 POTENTIAL.
FT TRANSMEM 43 63 POTENTIAL.
FT TRANSMEM 68 88 POTENTIAL.

```

```

FT TRANSMEM 89 109 POTENTIAL.
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 160 180 POTENTIAL.
FT TRANSMEM 210 230 POTENTIAL.
FT TRANSMEM 238 258 POTENTIAL.
FT TRANSMEM 259 279 POTENTIAL.
FT TRANSMEM 286 306 POTENTIAL.
SQ SEQUENCE 330 AA; 34456 MW; 87CDB0958FB81C1E CRC64;

Query Match 5.1%; Score 7; DB 1; Length 330;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 12 LAIITL 18
Db 88 LAIITL 94

RESULT 27
CDC2_CABEL STANDARD; PRT; 332 AA.
AC P34556;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cell division control protein 2 homolog (EC 2.7.1.-) (p34 protein kinase)
GN NCC-1 OR T05G5.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kerhaw J., Kirsten J., Latster N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Sims M., Smailson N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX Ferraz C., Thierry-Mieg D., le Peuch C.J.;
RL Submitted (SEP-1992) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95131956; PubMed=7830726;
RA Mori H., Palmer R.E., Sternberg P.W.;
RT "The identification of a Caenorhabditis elegans homolog of p34cdc2
RT kinase."
RL Mol. Genet. 245:781-786(1994).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99225469; PubMed=10207147;
RA Boxem M., Srinivasan D.G., van den Heuvel S.;
RT "The Caenorhabditis elegans gene ncc-1 encodes a cdc2-related kinase
RT required for M phase in meiotic and mitotic cell divisions, but not
RT for S phase."
RL Development 126:2227-2239(1999).
CC -1- FUNCTION: PLAYS A KEY ROLE IN THE CONTROL OF THE EUKARYOTIC
CC CELL CYCLE. IT IS REQUIRED IN HIGHER CELLS FOR ENTRY INTO S-PHASE

```

```

CC AND MITOSIS. P34 IS A COMPONENT OF THE KINASE COMPLEX THAT
CC PHOSPHORYLATES THE REPETITIVE CARBOXYL-TERMINUS OF RNA
CC POLYMERASE II (BY SIMILARITY).
CC -1- ENZYME REGULATION: PHOSPHORYLATION INACTIVATES THE ENZYME
CC (BY SIMILARITY).
CC -1- SUBUNIT: FORMS A STABLE BUT NON-COVALENT COMPLEX WITH A REGULATORY
CC SUBUNIT AND WITH A CYCLIN.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC CDC2/CDKX SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z27079; CAA81590.1; -
DR EMBL; X68384; CAA48455.1; -
DR EMBL; S75262; AAC60520.1; -
DR EMBL; AF129109; AAD37119.1; -
DR PIR; S41003; S41003.
DR HSSP; P24941; ICKP.
DR WormPep; T05G5.3; CE00315.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_Thr_kinase.
DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
KM Transferase, Serine/threonine-protein kinase; ATP-binding;
KM Cell cycle; Cell division; Mitosis; Phosphorylation.
FT DOMAIN 22 312
FT NP_BIND 28 36
FT BINDING 51 51
FT ACT_SITE 146 146
FT MOD_RES 32 32
FT MOD_RES 33 33
FT COMFLUCT 176 176
SQ SEQUENCE 332 AA; 38295 MW; B6297E92949C8206 CRC64;

Query Match 5.1%; Score 7; DB 1; Length 332;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 64 FRIFRVL 70
Db 232 FRIFRVL 238

RESULT 28
PEXC_CABEL STANDARD; PRT; 359 AA.
AC Q19189;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative peroxisome assembly protein 12 (peroxin-12).
GN P08B12.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX Dobson R.;
RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: Required for protein import into peroxisomes (By
CC similarity).

```

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Peroxisomal (potential).  
 CC -1- SIMILARITY: BELONGS TO THE PEROXIN 12 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL; Z68104; CA92113.1; -.  
 DR PIR; T20575; T20575.  
 DR WormPep; F08B12.2; CE03163.  
 DR InterPro; IPR006845; Pex2\_Pex12.  
 DR InterPro; IPR001841; Znf\_Fing.  
 DR Pfam; PF04757; Pex2\_Pex12; 1.  
 DR SMART; SM00184; RING; 1.  
 DR Hypothetical protein; Transmembrane; Peroxisome.  
 KW TRANSMEM 243 265 POTENTIAL.  
 FT SEQUENCE 359 AA; 41289 MW; 639057240A1BA93C CRC64;  
 SO  
 Query Match 5.1%; Score 7; DB 1; Length 359;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 28 KTFPLSVH 34  
 DB 55 KTFPLSVH 61  
 RESULT 29  
 CYB\_ANISC STANDARD; PRT; 371 AA.  
 AC 048014; (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Cytochrome b.  
 GN MT-CYB OR COB OR CYTB.  
 OS Anilius scyrale (Red pipe snake).  
 OG Mitochondrion.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Lepidodactylia; Squamata; Scleroglossa; Serpentes; Hemophidia; Aniliidae; Anilius.  
 OX NCBI\_TaxID=51844;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Campbell B.N.;  
 RT "Hic Sunt Serpentes - molecular phylogenetics and the Boidae (Serpentes: Boidae)".  
 RL Thesis (1997), Queen's University / Kingston, Canada.  
 CC -1- FUNCTION: Component of the ubiquinol-cytochrome c reductase complex (complex III or cytochrome b-c1 complex), which is a respiratory chain that generates an electrochemical potential coupled to ATP synthesis (By similarity).  
 CC -1- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or b562) is low-potential and absorbs at about 562, and heme 2 (or BH or b566) is high-potential and absorbs at about 566 (By similarity).  
 CC -1- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b, cytochrome c1 and the Rieske protein (By similarity).  
 CC -1- SIMILARITY: Belongs to the cytochrome b family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----

DR EMBL; U69738; AAD13429.1; -.  
 DR InterPro; IPR005798; Cytb\_b6\_C.  
 DR InterPro; IPR005797; Cytb\_b6\_N.  
 DR Pfam; PF00032; cytochrome b\_C; 1.  
 DR Pfam; PF00033; cytochrome b\_N; 1.  
 DR PROSITE; PS00193; CYTOCHROME B\_OO; 1.  
 DR PROSITE; PS00192; CYTOCHROME B\_HEME; FALSE NEG.  
 KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane; Heme.  
 FT METAL 75 75  
 FT METAL 89 89  
 FT METAL 174 174  
 FT METAL 188 188  
 FT SEQUENCE 371 AA; 41644 MW; 000DCD8427B16C55 CRC64;  
 SO  
 Query Match 5.1%; Score 7; DB 1; Length 371;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 14 ILTLTMA 20  
 DB 111 ILTLTMA 117  
 RESULT 30  
 GSPF\_KLEPN STANDARD; PRT; 361 AA.  
 AC P15745;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE General secretion pathway protein P (PulL) pulP.  
 GN PULF.  
 OS Klebsiella pneumoniae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Klebsiella.  
 OX NCBI\_TaxID=573;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Pugsley A.P., D'Entfer C., Reyes I., Wandersman C.;  
 RL Submitted (MAR-1990) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLLOCATION OF PULULINASE.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane (probable).  
 CC -1- SIMILARITY: BELONGS TO THE PULF/OUTF/EXEF/XPSF/XGPS FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL; M32613; AAA25128.1; -.  
 DR PIR; S20035; S20035.  
 DR InterPro; IPR003004; Bac\_GSPF.  
 DR InterPro; IPR001992; Bac\_sec\_systII.  
 DR Pfam; PF00482; GSPF1; 1.  
 DR PRINTS; PR00812; BCTERIALGSPF.  
 DR PROSITE; PS00874; T2SP\_F; 1.  
 KW Transport; Transmembrane; Inner membrane.  
 FT TRANSMEM 145 165  
 FT TRANSMEM 198 218  
 FT TRANSMEM 353 373  
 FT SEQUENCE 381 AA; 41872 MW; 01347C0994B29933 CRC64;  
 SO  
 Query Match 5.1%; Score 7; DB 1; Length 381;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 LLLAILL 16  
 DB 203 LLLAILL 209

RESULT 31  
 TNR3 MOUSE STANDARD, PRT, 415 AA.

AC P50264;  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 42, Last annotation update)  
 DE Tumor necrosis factor receptor superfamily member 3 precursor  
 DE (lymphotoxin-beta receptor).  
 GN LTRB OR TNFRSF3 OR TNFR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CVB; TISSUE=Lung;  
 RX MEDLINE=96072804; PubMed=7594541;  
 RA Browning J.L., Ware C.F.;  
 RT "Mouse lymphotoxin-beta receptor. Molecular genetics, ligand binding,  
 RT and expression."; Immunol. 155:5280-5288(1995).  
 RL J. Immunol. 155:5280-5288(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96163885; PubMed=8586432;  
 RA Nakamura T., Tashiro K., Nazarea M., Nakano T., Sasayama S.,  
 RA Honjo T.;  
 RT "The murine lymphotoxin-beta receptor cDNA: isolation by the signal  
 RT sequence trap and chromosomal mapping."; Genomics 30:312-319(1995).  
 RL [3]  
 RP INTERACTION WITH TRAF5.  
 RC STRAIN=Balb/c;  
 RX MEDLINE=96278943; PubMed=8663299;  
 RA Nakano H., Oshima H., Chung W., Williams-Abbott L., Ware C.F.,  
 RA Yagita H., Okumura K.;  
 RT "TRAF5, an activator of NF-kappaB and putative signal transducer for  
 RT the lymphotoxin-beta receptor."; J. Biol. Chem. 271:14661-14664(1996).  
 RL J. Biol. Chem. 271:14661-14664(1996).  
 CC -1- FUNCTION: Receptor for the heterotrimeric lymphotoxin containing  
 CC LVA and LTB, and for TNFSF14/LIGHT. Promotes apoptosis via TRAF3  
 CC and TRAF5. May play a role in the development of lymphoid organs  
 CC (by similarity).  
 CC -1- SUBUNIT: Self-associates (by similarity). Associates with TRAF5.  
 CC Associates with TRAF3 and TRAF4 (by similarity).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: Contains 4 TNFR-Cys repeats.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, U29173; AAA68864.1; -;  
 CC EMBL, L38423; AAB00846.1; -;  
 CC EMBL, U30798; AAA81334.1; -;  
 CC HSSP, O14763, IDOG.  
 CC MGD; MGI:104875; Lbdr.  
 CC InterPro: IPR001368; TNFR\_C6.  
 CC Pfam; PF00020; TNFR\_C6; 3.  
 CC SMART; SM00208; TNFR; 3.  
 CC PROSITE; PS00652; TNFR\_NGFR\_1; 2.  
 CC PROSITE; PS00650; TNFR\_NGFR\_2; 3.  
 CC Receptor; Apoptosis; Transmembrane; Glycoprotein; Repeat; Signal.

FT SIGNAL 1 30  
 FT CHAIN 31 415  
 FT DOMAIN 31 223  
 FT TRANSMEM 224 244  
 FT DOMAIN 245 415  
 FT REPEAT 42 81  
 FT REPEAT 82 124  
 FT REPEAT 125 170  
 FT REPEAT 171 213  
 FT DISULFID 43 58  
 FT DISULFID 59 72  
 FT DISULFID 62 80  
 FT DISULFID 83 98  
 FT DISULFID 101 116  
 FT DISULFID 104 124  
 FT DISULFID 126 132  
 FT DISULFID 139 150  
 FT DISULFID 142 169  
 FT DISULFID 172 187  
 FT CARBOHYD 40 40  
 FT CARBOHYD 179 179  
 SQ SEQUENCE 415 AA; 44956 MW; 29B326A56A8F6E1 CRC64;

Query Match 5.14; Score 7; DB 1; Length 415;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 LLLAILL 16  
 DB 222 LLLAILL 228

RESULT 32  
 TBBI SOYBN STANDARD, PRT, 445 AA.

AC P12459;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Tubulin beta-1 chain.  
 GN TUBB1.  
 OS Glycine max (Soybean).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
 OX NCBI\_TaxID=3847;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Gulltman M.J., Ma D.-P., Barker R.F., Bustos M.M., Cyr R.J.,  
 RA Yadegari R., Foeke D.E.;  
 RT "The isolation, characterization and sequence of two divergent  
 RT beta-tubulin genes from soybean (Glycine max L.).";  
 RL Plant Mol. Biol. 10:171-184(1987).  
 CC -1- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES. IT  
 CC BINDS TWO MOLES OF GTP. ONE AT AN EXCHANGEABLE SITE ON THE BETA  
 CC CHAIN AND ONE AT A NONEXCHANGEABLE SITE ON THE ALPHA-CHAIN.  
 CC -1- SUBUNIT: Dimer of alpha and beta chains.  
 CC -1- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL, M21296; AAA34009.1; -;  
 CC InterPro: IPR002453; Beta\_tubulin.  
 CC InterPro: IPR000217; Tubulin.  
 CC InterPro: IPR003008; Tubulin\_Fte2.  
 CC Pfam; PF00091; tubulin; 1.

DR Pfam: PF03953; tubulin C. 1.  
 DR PRINTS; PR01161; TUBULIN.  
 DR PROSITE; PS00227; TUBULIN. 1.  
 DR PROSITE; PS00228; TUBULIN B AUTOREG. 1.  
 DR Microcubules; GTP-binding; Multigene family.  
 FT NP BIND 140 146 GTP (POTENTIAL).  
 FT SEQUENCE 445 AA; 49991 MW; B68E4260D9AELDC CRC64;  
 SQ

Query Match 5.1%; Score 7; DB 1; Length 445;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 FSVFPAV 122  
 DB 167 FSVFPAV 173

RESULT 33  
 KCS1\_MOUSE STANDARD; PRT; 497 AA.  
 ID KCS1\_MOUSE  
 AC 035173;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Potassium voltage-gated channel subfamily S member 1 (Potassium channel Kv9.1) (Delayed-rectifier K+ channel alpha subunit 1).  
 GN KCNS1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND FUNCTION.  
 RC TISSUE=Brain;  
 RX MEDLINE=97450962; PubMed=9305895;  
 RA Salinas M., Duprat F., Heurteaux C., Hugnot J.-P., Lazdunski M.;  
 RT "New modulatory alpha subunits for mammalian Shab K+ channels."; J. Biol. Chem. 272:24371-24379(1997).  
 RL J. Biol. Chem. 272:24371-24379(1997).  
 CC -1- FUNCTION: Potassium channel subunit. Modulates channel activity and reduces the ion flow.  
 CC -1- SUBUNIT: Heteromultimer with KCNB1 and with KCNB2. Does not form homomultimers. Might also bind to other channel proteins.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. May not reach the plasma membrane but remain in an intracellular compartment in the absence of KCNB1.  
 CC -1- TISSUE SPECIFICITY: Detected in brain, but not in the other tissues tested. The highest levels of expression are in olfactory bulb, cerebral cortex, hippocampus, habenula, basolateral amygdaloid nuclei and cerebellum.  
 CC -1- DOMAIN: The segment S4 is probably the voltage-sensor and is characterized by a series of positively charged amino acids at every third position.  
 CC -1- SIMILARITY: BELONGS TO THE POTASSIUM CHANNEL FAMILY. S SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk).  
 CC -----  
 DR EMBL; AF008573; AAB72050.1; -.  
 DR HSBP; O54397; 1BL8.  
 DR MGD; MGI:1197019; Kcns1.  
 DR InterPro; IPR000210; BTB\_POZ.  
 DR InterPro; IPR005821; Ion\_trans.  
 DR InterPro; IPR001622; K+channel\_pore.  
 DR InterPro; IPR003091; K\_channel.  
 DR InterPro; IPR003131; K\_tetra.  
 DR InterPro; IPR003971; Kv9\_channel.  
 DR InterPro; IPR003968; Kv\_channel.  
 DR InterPro; IPR003820; M+channel\_nlg.

DR Pfam: PF00520; Ion\_trans; 1.  
 DR Pfam: PF02214; K\_tetra; 1.  
 DR PRINTS; PR00169; KCHANNEL.  
 DR PRINTS; PR01494; KV9CHANNEL.  
 DR PRINTS; PR01491; KVCHANNEL.  
 DR SMART; SM00225; BTB; 1.  
 DR Transport; Ion transport; Ionic channel; Voltage-gated channel;  
 KW Potassium channel; Potassium; Potassium transport; Transmembrane;  
 KM Multigene family.  
 KW DOMAIN 1  
 FT TRANSMEM 189  
 FT TRANSMEM 190 210 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 244 264 SEGMENT S1 (POTENTIAL).  
 FT TRANSMEM 265 275 SEGMENT S2 (POTENTIAL).  
 FT TRANSMEM 276 296 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 312 332 SEGMENT S3 (POTENTIAL).  
 FT TRANSMEM 333 347 SEGMENT S4 (POTENTIAL).  
 FT TRANSMEM 348 368 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 380 400 SEGMENT S5 (POTENTIAL).  
 FT TRANSMEM 408 428 SEGMENT H5 (PORE-FORMING) (POTENTIAL).  
 FT TRANSMEM 429 497 SEGMENT S6 (POTENTIAL).  
 FT DOMAIN 219 228 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN POLY-ALA.  
 SQ SEQUENCE 497 AA; 54903 MW; F732A5D4B65BBE CRC64;  
 SQ

Query Match 5.1%; Score 7; DB 1; Length 497;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 RIRRVLK 71  
 DB 320 RIRRVLK 326

RESULT 34  
 KCS1\_RAT STANDARD; PRT; 497 AA.  
 ID KCS1\_RAT  
 AC 088758;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Potassium voltage-gated channel subfamily S member 1 (Potassium channel Kv9.1) (Delayed-rectifier K+ channel alpha subunit 1).  
 GN KCNS1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=98369652; PubMed=9704029;  
 RA Stocker M., Kerscheneitner D.;  
 RT "Cloning and tissue distribution of two new potassium channel alpha-subunits cloned from rat brain."; Biochem. Biophys. Res. Commun. 248:927-934(1998).  
 RL Biochem. Biophys. Res. Commun. 248:927-934(1998).  
 CC -1- FUNCTION: Potassium channel subunit. Modulates channel activity and reduces the ion flow (By similarity).  
 CC -1- SUBUNIT: Heteromultimer with KCNB1 and with KCNB2. Does not form homomultimers. Might also bind to other channel proteins (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. May not reach the plasma membrane but remain in an intracellular compartment in the absence of KCNB1 (By similarity).  
 CC -1- TISSUE SPECIFICITY: Highly expressed in brain, but not in the other tissues tested.  
 CC -1- DOMAIN: The segment S4 is probably the voltage-sensor and is characterized by a series of positively charged amino acids at every third position.  
 CC -1- SIMILARITY: BELONGS TO THE POTASSIUM CHANNEL FAMILY. S SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL; Y17606; CAAT76804.1; -.

DR HSSB; Q54397; 1BL8.

DR InterPro; IPR000210; BTB\_POZ.

DR InterPro; IPR005821; Ion trans.

DR InterPro; IPR001622; K+channel\_pore.

DR InterPro; IPR003091; K channel.

DR InterPro; IPR003311; K\_tetra.

DR InterPro; IPR003971; Kv9\_channel.

DR InterPro; IPR003968; Kv\_channel.

DR InterPro; IPR005820; M+channel\_nlg.

DR Pfam; PF00520; Ion trans; 1.

DR Pfam; PF02214; K\_tetra; 1.

DR PRINTS; PR00169; KCHANNEL.

DR PRINTS; PR01494; KV9CHANNEL.

DR PRINTS; PR01491; KVCHANNEL.

DR SMART; SM00225; BTB; 1.

DR Transport; Ion transport; Ionic channel; Voltage-gated channel;  
 KW Potassium channel; Potassium; Potassium transport; Transmembrane;  
 KM Multigene family.

FT DOMAIN 1 189 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 190 210 SEGMENT S1 (POTENTIAL).

FT TRANSMEM 244 264 SEGMENT S2 (POTENTIAL).

FT DOMAIN 265 275 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 276 296 SEGMENT S3 (POTENTIAL).

FT TRANSMEM 312 332 SEGMENT S4 (POTENTIAL).

FT DOMAIN 333 347 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 348 368 SEGMENT S5 (POTENTIAL).

FT DOMAIN 380 400 SEGMENT S5 (PORE-FORMING) (POTENTIAL).

FT TRANSMEM 408 428 SEGMENT S6 (POTENTIAL).

FT DOMAIN 429 497 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 219 228 POLY-ALA.

SO SEQUENCE 497 AA; 54914 MW; 268AE4D56051F7A0 CRC64;

Query Match 5.1%; Score 7; DB 1; Length 497;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 65 RIFRYLK 71  
 |||||  
 Db 320 RIFRYLK 326

RESULT 35

CCSI\_HUMAN STANDARD; PRT; 526 AA.

ID CCSI\_HUMAN

AC O96K63; O43652;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Potassium voltage-gated channel subfamily S member 1 (Potassium  
 DE channel Kv9.1) (Delayed-rectifier K+ channel alpha subunit 1).  
 GN KCSI1.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain, and Lens epithelium;  
 RX MEDLINE=99413882; PubMed=10464328;  
 RA Shepard A.R., Rae J.L.;  
 RT "Electrically silent potassium channel subunits from human lens  
 RT epithelium.";  
 RL Am. J. Physiol. 277:C412-C424(1999).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=21638749; PubMed=11780052;  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,

RA Bailey J., Bartlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
 RA Bessley O.P., Bird C.P., Blakey S.B., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burrill W., Butler A.P., Carder C., Carter N.P.,  
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.S., Collier R.E., Connor R., Cordy N.R.,  
 RA Coulson A., Coville G.J., Deedman R., Dhami P., Dunn M.,  
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Giffham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.M., Lawlor S.,  
 RA Leivasalho M.H., Leversha M., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McConnachie I.J., McKay K., McMurray A.A.,  
 RA Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhen R., Sims S.,  
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sultoni J.E.,  
 RA Swann M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.

RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).

CC -1- FUNCTION: Potassium channel subunit. Modulates channel activity  
 CC and reduces the ion flow (By similarity).

CC -1- SUBUNIT: Heteromultimer with KCNB1 and with KCNB2. Does not form  
 CC homomultimers. Might also bind to other channel proteins (By  
 CC similarity).

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. May not reach the  
 CC plasma membrane but remain in an intracellular compartment in the  
 CC absence of KCNB1.

CC -1- TISSUE SPECIFICITY: Detected in all tissues tested with the  
 CC exception of skeletal muscle. Highly expressed in adult and fetal  
 CC brain, fetal kidney and lung, and adult prostate and testis.

CC -1- DOMAIN: The segment S4 is probably the voltage-sensor and is  
 CC characterized by a series of positively charged amino acids at  
 CC every third position.

CC -1- SIMILARITY: BELONGS TO THE POTASSIUM CHANNEL FAMILY. S SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF043473; AACJ3165.2; -.  
 DR EMBL; Z93016; CAB07507.2; -.  
 DR Genew; HGNC:6300; KCSI1.  
 DR MIM; 602905; -.

DR InterPro; IPR000210; BTB\_POZ.  
 DR InterPro; IPR005821; Ion trans.  
 DR InterPro; IPR001622; K+channel\_pore.  
 DR InterPro; IPR003091; K channel.  
 DR InterPro; IPR003311; K\_tetra.  
 DR InterPro; IPR003971; Kv9\_channel.  
 DR InterPro; IPR003968; Kv\_channel.  
 DR InterPro; IPR005820; M+channel\_nlg.  
 DR Pfam; PF00520; Ion trans; 1.  
 DR Pfam; PF02214; K\_tetra; 1.  
 DR PRINTS; PR00169; KCHANNEL.  
 DR PRINTS; PR01494; KV9CHANNEL.  
 DR PRINTS; PR01491; KVCHANNEL.  
 DR SMART; SM00225; BTB; 1.  
 DR Transport; Ion transport; Ionic channel; Voltage-gated channel;  
 KW Potassium channel; Potassium; Potassium transport; Transmembrane;  
 KM Multigene family.

FT DOMAIN 1 221 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 222 242 SEGMENT S1 (POTENTIAL).

FT TRANSMEM 275 295 SEGMENT S2 (POTENTIAL).

```

FT DOMAIN 296 306 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 307 327 SEGMENT S3 (POTENTIAL).
FT TRANSMEM 340 360 SEGMENT S4 (POTENTIAL).
FT DOMAIN 361 375 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 376 396 SEGMENT S5 (POTENTIAL).
FT DOMAIN 409 429 SEGMENT H5 (POTENTIAL).
FT TRANSMEM 437 457 SEGMENT S6 (POTENTIAL).
FT DOMAIN 458 526 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 250 259 POLY-ALA.
SQ SEQUENCE 526 AA; 58371 MW; 00F0E415B3A68C91 CRC64;

Query Match 5.1%; Score 7; DB 1; Length 526;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 65 RIFRVLK 71
Db 348 RIFRVLK 354

RESULT 36
MTLR STRMU STANDARD; PRT; 650 AA.
ID MTLR STRMU STANDARD; PRT; 650 AA.
AC 002435;
DT 01-JUL-1993 (Rel. 26, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative transcriptional regulator mtlr.
GN MTLR OR SMU.1184C.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OC NCBI_TaxID=1309;
RN NCBI_TaxID=1309;
RP SEQUENCE FROM N.A. AND FUNCTION.
RC STRAIN=UA130 / Serotype C;
RX MEDLINE=20340960; PubMed=10878121;
RA Honeyman A.L., Curtiss R. III;
RT "The mannitol-specific enzyme II (mtIIA) gene and the mtlr gene of the
RT PTS of Streptococcus mutans."
RL Microbiology 146:1565-1572(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=UA159 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Aidic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Ian S., Qian Y.,
RA Li S., Zhu H., Najjar P., Lai H., White J., Roe B.A., Perretti J.J.;
RT "genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen."
RT Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
RN [3]
RP SEQUENCE OF 525-650 FROM N.A.
RC STRAIN=UA130 / Serotype C;
RX MEDLINE=92348013; PubMed=1322373;
RA Honeyman A.L., Curtiss R. III;
RT "Isolation, characterization, and nucleotide sequence of the
RT Streptococcus mutans mannitol-phosphate dehydrogenase gene and the
RT mannitol-specific factor III gene of the phosphoenolpyruvate
RT phosphotransferase system."
RL Infect. Immun. 60:3369-3375(1992).
RN [4]
RP FUNCTION: NOT NECESSARY FOR MANNITOL UTILIZATION. MAY BE INVOLVED
RP IN REGULATION OF THE MANNITOL PHOSPHOENOLPYRUVATE-DEPENDENT SUGAR
RP PHOSPHOTRANSFERASE SYSTEM (PTS).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).

```

```

DR EMBL; AF210133; AAA26940.2; -.
DR EMBL; AB014955; AAN58874.1; -.
DR PIR; A44798; A44798.
KW Transcription regulation; Complete proteome.
FT CONFLICT 361 363 KEP -> RES (IN REF. 1).
FT CONFLICT 441 441 V -> A (IN REF. 1).
FT CONFLICT 447 447 M -> I (IN REF. 1).
FT CONFLICT 447 447 K -> R (IN REF. 1).
FT CONFLICT 482 482 C -> R (IN REF. 1 AND 3).
FT CONFLICT 539 539 H -> R (IN REF. 1 AND 3).
FT CONFLICT 551 551 L -> P (IN REF. 1 AND 3).
FT CONFLICT 620 620 L -> P (IN REF. 1 AND 3).
SQ SEQUENCE 650 AA; 75337 MW; 02F1229435B7C97A CRC64;

Query Match 5.1%; Score 7; DB 1; Length 650;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 LVALILT 17
Db 151 LVALILT 157

RESULT 37
C180 HUMAN STANDARD; PRT; 661 AA.
ID C180 HUMAN STANDARD; PRT; 661 AA.
AC 099467;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE C0180 antigen precursor (lymphocyte antigen 64) (Radioprotective 105
DE kDa protein).
GN LY64 OR RP105.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=B-cell lymphoma;
RX MEDLINE=97131508; PubMed=8975706;
RA Miura Y., Miyake K., Yamashita Y., Shimazu R., Copeland N.G.,
RA Gilbert D.J., Jenkins N.A., Inazawa J., Abe T., Kimoto M.;
RT "Molecular cloning of a human RP105 homologue and chromosomal
RT localization of the mouse and human RP105 genes (Ly64 and LY64).";
RL Genomics 38:299-304(1996).
RN [2]
RP INTERACTION WITH MD-1.
RX MEDLINE=98438337; PubMed=9763566;
RA Miura Y., Shimazu R., Miyake K., Akashi S., Ogata H., Yamashita Y.,
RA Natsawa Y., Kimoto M.;
RT "RP105 is associated with MD-1 and transmits an activation signal in
RT human B cells."
RL Blood 92:2815-2822(1998).
RN [3]
RP FUNCTION: May cooperate with MD-1 and TLR4 to mediate the innate
RP immune response to bacterial lipopolysaccharide (LPS) in B cells.
RP Leads to NF-kappa-B activation. Also involved in the life/death
RP decision of B cells (By similarity).
CC -1- SUBUNIT: Binds to MD-1.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Plasma membrane.
CC -1- TISSUE SPECIFICITY: Expressed mainly on mature peripheral B
CC cells. Detected in spleen, lymph node and appendix. Not detected
CC in pre-B and -T cells.
CC -1- DISEASE: Decreased expression of LY64 has been shown in systemic
CC lupus erythematosus (SLE). The loss of LY64 is associated with B-
CC cell activation and increased disease activity in SLE patients.
CC -1- SIMILARITY: BELONGS TO THE TOLL-LIKE RECEPTOR FAMILY.
CC -1- SIMILARITY: Contains 17 leucine-rich (LRR) repeats.
CC WWW="http://www.ncbi.nlm.nih.gov/prov/guide/340502642_g.htm".
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.iesb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

-----  
 DR EMBL: D83597; BAI2019.1; -  
 DR Genew; HGNC:6726; LY64.  
 DR MIM; 602226; -  
 DR GO; GO:0005515; F:protein binding activity; TAS.  
 DR InterPro; IPR01611; LRR.  
 DR InterPro; IPR00483; LRR\_Chern.  
 DR Pfam; PF00560; LRR\_9.  
 DR Pfam; PF01463; LRRCT; 1.  
 DR SMART; SM00082; LRRCT; 1.  
 KW Receptor; Immune response; Inflammatory response; Signal; Antigen;  
 KW Transmembrane; Repeat; Leucine-rich repeat; Glycoprotein.  
 FT SIGNAL 1 23  
 FT CHAIN 24 661  
 FT DOMAIN 24 626  
 FT TRANSMEM 627 650  
 FT DOMAIN 651 661  
 FT REPEAT 51 75  
 FT REPEAT 76 99  
 FT REPEAT 101 123  
 FT REPEAT 148 172  
 FT REPEAT 174 195  
 FT REPEAT 198 221  
 FT REPEAT 223 296  
 FT REPEAT 297 320  
 FT REPEAT 323 346  
 FT REPEAT 353 369  
 FT REPEAT 370 392  
 FT REPEAT 395 418  
 FT REPEAT 420 443  
 FT REPEAT 468 494  
 FT REPEAT 496 518  
 FT REPEAT 519 540  
 FT REPEAT 541 565  
 FT CARBOHYD 34 34  
 FT CARBOHYD 53 53  
 FT CARBOHYD 70 70  
 FT CARBOHYD 78 78  
 FT CARBOHYD 201 201  
 FT CARBOHYD 234 234  
 FT CARBOHYD 244 244  
 FT CARBOHYD 394 394  
 FT CARBOHYD 402 402  
 FT CARBOHYD 451 451  
 FT CARBOHYD 573 573  
 SQ SEQUENCE 661 AA; 74164 MW; C08F2A5E8DE5A607 CRC64;

Query Match 5.1%; Score 7; DB 1; Length 661;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 LLLAIIIL 16  
 Db 640 LLLAIIIL 646

RESULT 38  
 PRML\_DROME STANDARD; PRT; 1013 AA.  
 AC P82295; Q95RKL; Q9VZV0;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Prominin-like protein.  
 GN CG7740.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryo; N.A.  
 RA Cordell D., Roper K., Hutterer W.B.;  
 RT "Sequence analysis of a Drosophila ortholog of mouse prominin and  
 RL human prominin(mouse)-like 1.";  
 RN Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amaratunga P.G., Scherer S.E., Li P.W., Hinkley R.A., Gale R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Miller B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris W.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hosteln D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,  
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshref A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheefel F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RN Science 287:2185-2195(2000).  
 [3]  
 RP REVISIONS.  
 RC STRAIN=Berkley;  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mangall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Betencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
 RL systematic review.";  
 RN Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).  
 [4]  
 RP SEQUENCE OF 482-1013 FROM N.A.  
 RC STRAIN=Berkley; TISSUE=Embryo, and Ovary;  
 RX MEDLINE=22426066; PubMed=12537569;  
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
 RA George R.A., Guatin H., Krommiller B., Pacle J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celniker S.E.;  
 RT "A Drosophila full-length cDNA resource.";  
 RN Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).



FT CARBOHYD 347 347 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 433 433 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 3396 AA; 379558 MW; C53E75F3E424367D CRC64;

Query Match 5.1%; Score 7; DB 1; Length 3396;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 105 EREHKO 111  
 Db 2929 EREHKO 2935

RESULT 40  
 ATP8\_PARLI  
 ID ATP8\_PARLI STANDARD; PRT; 54 AA.

AC P1267;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE ATP synthase protein 8 (EC 3.6.3.14) (ATPase subunit 8) (A6L).  
 GN ATP8 OR ATP8.  
 OS Paracentrotus lividus (Common sea urchin).  
 OC Paracentrotus lividus; Echinodermata; Eleutherozoa; Echinozoa;  
 OC Eukaryota; Metazoa; Echinodermata; Echinozoa; Echinodermata;  
 OC Echinodermata; Echinodermata; Echinodermata; Echinodermata;  
 OC Paracentrotus.  
 CC Echinodermata.  
 CC NCBI\_TaxId=7656;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=89291831; PubMed=2544576;  
 RA Cantatore P., Roberti M., Rainaldi G., Gadaleira M.N., Saccoccia C.;  
 RT "The complete nucleotide sequence, gene organization, and genetic  
 code of the mitochondrial genome of Paracentrotus lividus.";  
 RL J. Biol. Chem. 264:10965-10975 (1989).  
 CC -1- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT  
 (CF(0) SUBUNIT) OF THE MITOCHONDRIAL ATPASE COMPLEX.  
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +  
 H(+) (Out).  
 CC -1- SUBCELLULAR LOCATION: Membrane-bound.  
 CC -1- SIMILARITY: BELONGS TO THE ATPASE PROTEIN 8 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; J04815; AAA68138.1; -.  
 DR PIR; F34284; F34284.  
 DR InterPro; IPR001421; ATPase8\_mit.  
 DR Pfam; PF00895; ATP-synt 8; 1.  
 KM Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.  
 FT TRANSMEM 8  
 FT SEQUENCE 54 AA; 6452 MW; 41B52C808607EA46 CRC64;

Query Match 4.4%; Score 6; DB 1; Length 54;  
 Best Local Similarity 100.0%; Pred. No. 32;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 ALLITL 18  
 Db 19 ALLITL 24

RESULT 41  
 HUNB\_APIME  
 ID HUNB\_APIME STANDARD; PRT; 69 AA.  
 AC P31504;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Hunchback protein (Fragment).  
 GN HB.

OS Apis mellifera (Honeybee).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;  
 OC Apidae; Apis.  
 CC NCBI\_TaxId=7460;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=9306327; PubMed=1438276;  
 RA Sommer R.J., Retzlaff M., Goerlich K., Sander K., Tautz D.;  
 RT "Evolutionary conservation pattern of zinc-finger domains of  
 Drosophila segmentation genes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:10782-10786 (1992).  
 CC -1- FUNCTION: GAP CLASS SEGMENTATION PROTEIN THAT CONTROLS DEVELOPMENT  
 OF HEAD STRUCTURES.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE HUNCHBACK FAMILY OF C2H2-TYPE ZINC-  
 FINGER PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; L01587; AAA27734.1; -.  
 DR InterPro; IPR007087; Znf C2H2.  
 DR Pfam; PF00096; zf-C2H2; 2.  
 DR SMART; SM00355; Znf C2H2; 2.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 1.  
 DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 2.  
 KM Developmental protein; Gap protein; Zinc-finger;  
 KM Metal-binding; DNA-binding; Repeat; Nuclear protein.  
 FT NOW TER 1  
 FT ZN\_FING <1 11 C2H2-TYPE.  
 FT ZN\_FING 17 39 C2H2-TYPE.  
 FT ZN\_FING 45 69 C2H2-TYPE.  
 FT NOW TER 69  
 SQ SEQUENCE 69 AA; 8230 MW; 61F7F5685261B3D4 CRC64;

Query Match 4.4%; Score 6; DB 1; Length 69;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 79 HLEVHL 84  
 Db 3 HLEVHL 8

RESULT 42  
 EXTS\_CLOAB  
 ID EXTS\_CLOAB STANDARD; PRT; 77 AA.  
 AC Q97HD1;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Probable exodeoxyribonuclease VII small subunit (EC 3.1.11.6)  
 GN (Exonuclease VII small subunit).  
 GN XSERB OR CAC2081.  
 OS Clostridium acetobutylicum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 CC Clostridium.  
 CC NCBI\_TaxId=1488;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
 RX MEDLINE=21359325; PubMed=11466286;  
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
 Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,

```

RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.,
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
CC -1- FUNCTION: Bidirectionally degrades single-stranded DNA into large
CC acid-insoluble oligonucleotides, which are then degraded further
CC into small acid-soluble oligonucleotides (By similarity).
CC -1- CATALYTIC ACTIVITY: Exonuclease cleavage in either 5' - to 3' -
CC or 3' - to 5' -direction to yield nucleoside 5'-phosphates.
CC -1- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE XSEB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB007710; AAK80040.1; -.
DR PIR: E97156; E97156.
DR HAMAP: MF_00337; -; 1.
DR InterPro: IPR003761; Exonuc VII S.
DR Pfam: PF02609; Exonuc VII S; 1.
DR ProDom: PD028235; Exonuc VII S; 1.
DR Hydrobase: Nuclease; Exonuclease; Complete proteome.
SQ SEQUENCE 77 AA; 8923 MW; FB895E7B24726603 CRC64;

Query Match 4.4%; Score 6; DB 1; Length 77;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 YKLINK 132
DB 47 YKLINK 52

RESULT 43
CYB CORME STANDARD; PRT; 79 AA.
ID CYB CORME STANDARD; PRT; 79 AA.
AC P16364;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cytochrome b (Fragment).
GN MTCYB OR COB OR CYTB.
OS Corcorax melanorhamphos (White-winged cichlid) (CROW).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauia; Aves; Neognathae; Passeriformes; Grallinidae;
OC Corcoracinae; Corcorax.
OX NCBI_TaxID=9145;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89345630; PubMed=2762322;
RA Kocher T.D., Thomas W.K., Meyer A., Edwards S.V., Paabo S.,
RA Villablanca F.X., Wilson A.C.;
RT "Dynamics of mitochondrial DNA evolution in animals: amplification
RT and sequencing with conserved primers.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:6196-6200(1989).
CC -1- FUNCTION: Component of the ubiquinol-cytochrome c reductase
CC complex (complex III or cytochrome b-c1 complex), which is a
CC respiratory chain that generates an electrochemical potential
CC coupled to ATP synthesis (By similarity).
CC -1- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or
CC b562) is low-potential and absorbs at about 562, and heme 2 (or BH
CC or b566) is high-potential and absorbs at about 566 (By
CC similarity).
CC -1- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,

```

```

CC cytochrome c1 and the Rieske protein (By similarity).
CC -1- SIMILARITY: Belongs to the cytochrome b family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M25690; AAA31708.1; -.
DR InterPro: IPR005798; Cytb_b6_C.
DR InterPro: IPR005797; Cytb_b6_N.
DR Pfam: PF00033; cytochrome_b_N; 1.
DR PROSITE: PS00193; CYTOCHROME_B_Q0; PARTIAL.
DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
KW Heme.
FT NON_TER 1 1
FT METAL 37 37 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 51 51 IRON 2 (HEME B566 AXIAL LIGAND).
FT NON_TER 79 79
SQ SEQUENCE 79 AA; 8931 MW; CA79ED7947A8FB44 CRC64;

Query Match 4.4%; Score 6; DB 1; Length 79;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 LITLMA 20
DB 74 LITLMA 79

RESULT 44
CYB POMSU STANDARD; PRT; 79 AA.
ID CYB POMSU STANDARD; PRT; 79 AA.
AC P16362;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cytochrome b (Fragment).
GN MTCYB OR COB OR CYTB.
OS Pomatosomus superciliosus (White-browed babler).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauia; Aves; Neognathae; Passeriformes; Timaliidae;
OC Pomatosomus.
OX NCBI_TaxID=9177;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89345630; PubMed=2762322;
RA Kocher T.D., Thomas W.K., Meyer A., Edwards S.V., Paabo S.,
RA Villablanca F.X., Wilson A.C.;
RT "Dynamics of mitochondrial DNA evolution in animals: amplification
RT and sequencing with conserved primers.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:6196-6200(1989).
CC -1- FUNCTION: Component of the ubiquinol-cytochrome c reductase
CC complex (complex III or cytochrome b-c1 complex), which is a
CC respiratory chain that generates an electrochemical potential
CC coupled to ATP synthesis (By similarity).
CC -1- COFACTOR: Binds two heme groups non-covalently. Heme 1 (or BL or
CC b562) is low-potential and absorbs at about 562, and heme 2 (or BH
CC or b566) is high-potential and absorbs at about 566 (By
CC similarity).
CC -1- SUBUNIT: The main subunits of complex b-c1 are: cytochrome b,
CC cytochrome c1 and the Rieske protein (By similarity).
CC -1- SIMILARITY: Belongs to the cytochrome b family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 DR EMBL; M25687; AAA32138.1; -  
 DR InterPro; IPR005798; Cyb\_b6\_C.  
 DR InterPro; IPR005797; Cyb\_b6\_N.  
 DR Pfam; PF00033; Cytochrome\_b\_N; 1.  
 DR PROSITE; PS00193; CYTOCHROME\_B\_CO; PARTIAL.  
 DR PROSITE; PS00192; CYTOCHROME\_B\_HEME; 1.  
 KM Electron transport; Mitochondrion; Respiratory chain; Transmembrane;  
 KM Heme.  
 FT NON\_TER 1 1  
 FT METAL 37 37 IRON 1 (HEME B562 AXIAL LIGAND).  
 FT METAL 51 51 IRON 2 (HEME B566 AXIAL LIGAND).  
 FT NON\_TER 79 79  
 SQ SEQUENCE 79 AA, 8883 MW, 9095B47A6210119 CRC64;

Query Match 4.4%; Score 6; DB 1; Length 79;  
 Best Local Similarity 100.0%; Pred. No. 45;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 15 LTTTMA 20  
 Db 74 LTTTMA 79

RESULT 45  
 DEFL\_CAVPO STANDARD; PRT; 93 AA.  
 AC P11478;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Neutrophil cationic peptide 1 precursor (Neutrophil defensin) (GNP)  
 OS (Corticostatic peptide GP-CSI) (CP-1).  
 OS Cavia porcellus (Guinea pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Hystriognathi; Caviidae; Cavia.  
 OX NCBI\_TaxID=10141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Hartley; TISSUE=Bone marrow;  
 RX MEDLINE=91192152; PubMed=2013325;  
 RA Nagaoka I., Someya A., Iwabuchi K., Yamashita T.;  
 RT "Characterization of cDNA clones encoding guinea pig neutrophil  
 cationic peptides.",  
 RL FEBS Lett. 280:287-291(1991).  
 RN [2]  
 RP SEQUENCE OF 63-93.  
 RC TISSUE=Peritoneal neutrophils;  
 RX MEDLINE=87306884; PubMed=3623703;  
 RA Seisted M.E., Harris S.S.L.;  
 RT "Purification, primary structure, and antimicrobial activities of a  
 guinea pig neutrophil defensin.",  
 RL Infect. Immun. 55:2281-2286(1987).  
 RN [3]  
 RP SEQUENCE OF 63-93.  
 RC TISSUE=Bone marrow;  
 RX MEDLINE=92062075; PubMed=1659400;  
 RA Hu J., Bennett H.P.J., Lazure C., Solomon S.;  
 RT "Isolation and characterization of corticostatic peptides from guinea  
 pig bone marrow.",  
 RL Biochem. Biophys. Res. Commun. 180:558-565(1991).  
 RN [4]  
 RP SEQUENCE OF 63-93.  
 RC TISSUE=Neutrophils;  
 RX MEDLINE=89307555; PubMed=2473036;  
 RA Yamashita T., Saito K.;  
 RT "Purification, primary structure, and biological activity of guinea  
 pig neutrophil cationic peptides.",  
 RL Infect. Immun. 57:2405-2408(1989).  
 CC -1- FUNCTION: THIS PEPTIDE HAS ANTIBIOTIC, ANTI-FUNGI AND ANTIVIRAL

CC ACTIVITY.  
 CC -1- SIMILARITY: BELONGS TO THE CORTICOSTATIN/DEFENSIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 DR EMBL; D14119; BAA03182.1; -  
 DR EMBL; X57705; CAA40880.1; ALT\_SEQ.  
 DR EMBL; D37971; BAA07189.1; -  
 DR PIR; S14314; S14314.  
 DR InterPro; IPR006081; Defensin\_alpha.  
 DR InterPro; IPR006080; Defensin\_mammal.  
 DR InterPro; IPR002366; Defensin\_propep.  
 DR Pfam; PF00879; Defensin\_propep; 1.  
 DR Pfam; PF00323; defensins; 1.  
 DR SMART; SM00048; DERSN; 1.  
 DR PROSITE; PS00269; DEFENSIN; 1.  
 KM Defensin; Antibiotic; Antiviral; Fungicide; Signal.  
 KM SIGNAL 1 19  
 FT PROPEP 20 62 POTENTIAL.  
 FT CHAIN 63 93 NEUTROPHIL CATIONIC PEPTIDE 1.  
 FT DISULFID 65 93 BY SIMILARITY.  
 FT DISULFID 67 82 BY SIMILARITY.  
 FT DISULFID 72 92 BY SIMILARITY.  
 SQ SEQUENCE 93 AA, 10479 MW, D85EAC5A1EB8A4B3 CRC64;

Query Match 4.4%; Score 6; DB 1; Length 93;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 15 LTTTMA 20  
 Db 12 LTTTMA 17

RESULT 46  
 DEFL\_CAVPO STANDARD; PRT; 93 AA.  
 AC P49112; Q9R025;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Neutrophil cationic peptide 2 precursor (CP-2) (GNCP) (GNCP-2).  
 OS Cavia porcellus (Guinea pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Hystriognathi; Caviidae; Cavia.  
 OX NCBI\_TaxID=10141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Hartley; TISSUE=Bone marrow;  
 RX MEDLINE=91192152; PubMed=2013325;  
 RA Nagaoka I., Someya A., Iwabuchi K., Yamashita T.;  
 RT "Characterization of cDNA clones encoding guinea pig neutrophil  
 cationic peptides.",  
 RL FEBS Lett. 280:287-291(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Hartley; TISSUE=Liver;  
 RX MEDLINE=92275076; PubMed=1592112;  
 RA Nagaoka I., Someya A., Iwabuchi K., Yamashita T.;  
 RT "Structure of the guinea pig neutrophil cationic peptide gene.",  
 RL FEBS Lett. 303:31-35(1992).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Hartley; TISSUE=Liver;  
 RX MEDLINE=94227245; PubMed=8173076;  
 RA Nagaoka I., Nonoguchi A., Yamashita T.;  
 RT "Cloning and characterization of the guinea pig neutrophil cationic

```

RT Peptide-1 and -2 genes.
RL DNA Seq. 4:123-128(1993).
CC -1- FUNCTION: THIS PEPTIDE HAS ANTI-BIOTIC, ANTI-FUNGI AND ANTI-VIRAL
CC ACTIVITY.
CC -1- SIMILARITY: BELONGS TO THE CORTICOSTATIN/DEFENSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X63676; CAA45215.1; -
DR EMBL; X63677; CAA45216.3; -
DR EMBL; D14121; BAA03184.1; -
DR EMBL; D14122; BAA03185.1; -
DR EMBL; D37974; BAA07192.1; -
DR EMBL; D37973; BAA07191.1; -
DR PIR; S21169; S21169.
DR InterPro; IPR006081; Defensin_alpha.
DR InterPro; IPR006080; Defensin_mammal.
DR InterPro; IPR002365; Defensin_propep.
DR Pfam; PF00879; Defensin_propep; 1.
DR SMART; SM00048; DEFSN; 1.
DR PROSITE; PS00269; DEFSIN; 1.
DR Defensin; Antibioc; Antiviral; Fungicide; Signal.
KM Defensin; Antibioc; Antiviral; Fungicide; Signal.
FT PROPEP 1 19 POTENTIAL.
FT CHAIN 20 62 NEUTROPHIL CATIONIC PEPTIDE 2.
FT DISULFID 65 93 BY SIMILARITY.
FT DISULFID 67 82 BY SIMILARITY.
FT DISULFID 72 92 BY SIMILARITY.
SQ SEQUENCE 93 AA; 10478 MW; 6D5AA0B1EBE8AB CRC64;

Query Match 4.4%; Score 6; DB 1; Length 93;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 LTLTMA 20
DB 12 LTLTMA 17

RESULT 47
Y874_METUA STANDARD; PRT; 93 AA.
ID Y874_METUA
AC Q58284;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein M0874.
GN M0874.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanodococcaceae; Methanodococcaceae.
NCBI_Taxid=2190;
OX NCBI_Taxid=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=868087;
RA Bult C.U., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Utechtack T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."

```

```

RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: TO THE C-TERMINAL OF B. SUBTILIS YDCN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67531; AAB98881.1; -
DR PIR; B64409; B64409.
DR TIGR; M0874; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 93 AA; 10860 MW; 07603737C0CC494D CRC64;

Query Match 4.4%; Score 6; DB 1; Length 93;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LTLAIL 15
DB 3 LTLAIL 8

RESULT 48
FTSB_ECOLI STANDARD; PRT; 103 AA.
ID FTSB_ECOLI
AC Q46894;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cell division protein ftsB.
GN FTSB OR B2748 OR Z4056 OR EC53602.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
NCBI_Taxid=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=9742617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=1126551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Renna D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kubara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).

```

```

RN [4]
RP CHARACTERIZATION.
RC STRAIN-K12 / MC4100.
RX MEDLINE=21980622; PubMed=11972052;
RA Buddemeier N., Judson N., Boyd D., Mekalanos J.J., Beckwith J.;
RT "yjbQ, a cell division protein in Escherichia coli and Vibrio
RT cholerae, localizes in codependent fashion with FtsL to the division
RT site";
RL Proc. Natl. Acad. Sci. U.S.A. 99:6316-6321(2002).
CC -1- FUNCTION: Required for the cell division process.
CC -1- SUBUNIT: May interact with FtsL.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane
CC (Potential). Colocalizes with FtsL to the division site.
CC -1- SIMILARITY: BELONGS TO THE FTSB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U29579; AAA6258.1; -.
DR EMBL; AE000358; AAC75790.1; -.
DR EMBL; AE005502; AAC57855.1; -.
DR EMBL; AP002562; BAB37025.1; -.
DR PIR; B91079; B91079.
DR PIR; C85924; C85924.
DR PIR; H65055; H65055.
DR EcoGene; EG1311; ftsB.
DR HAMAP; MF_00599; -.
DR InterPro; IPR007060; DivIC.
DR Pfam; PF04977; DivIC; 1.
DR Cell division; Transmembrane; Inner membrane; Coiled coil;
DR Complete proteome.
KW DOMAIN 1 3 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 4 21 POTENTIAL.
FT DOMAIN 22 103 PERIPLASMIC (POTENTIAL).
FT DOMAIN 29 75 COILED COIL (POTENTIAL).
SQ SEQUENCE 103 AA; 11622 MW; 9850A7EF637809D7 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 56;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLAAIL 15
Db 7 LLAAIL 12

RESULT 49
VNM SHVX STANDARD; PRT; 103 AA.
AC 004582;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE 11.2 kDa membrane protein (ORF 3).
OS Shalot virus X (SHVX).
CC Viruses; ssRNA positive-strand viruses, no DNA stage; Allexivirus.
CC NCBI_TaxID=31770;
CX [1]
RN SEQUENCE FROM N.A.
RA MEDLINE=3019008; PubMed=1339468;
RA Kanyuka K.V., Vishnichenko V.K., Levey K.E., Kondrikov D.Y.,
RA Ryabov B.V., Zavyetov S.K.;
RT "Nucleotide sequence of shalot virus X RNA reveals a 5'-proximal
RT ctarion closely related to those of potexviruses and a unique
RT arrangement of the 3'-proximal ctarions.";
RL J. Gen. Virol. 73:2553-2560(1992).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: TO ORF3 PROTEIN FROM OTHER POTEVIRUSES AND TO 12 kDa

```

```

CC PROTEIN FROM CARLAVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M97264; AAA47789.1; -.
DR PIR; J01736; J01736.
DR InterPro; IPR001896; Plant_vir_prot.
DR Pfam; PF01307; Plant_vir_prot; 1.
DR ProDom; PD001561; Plant_vir_prot; 1.
KW Transmembrane.
FT TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 71 91 POTENTIAL.
SQ SEQUENCE 103 AA; 11246 MW; FC2701EC78F9E4C1 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 56;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LAILLT 17
Db 83 LAILLT 88

RESULT 50
VNF ECOLI STANDARD; PRT; 113 AA.
ID VNF ECOLI
AC P76170;
DT 15-JUN-1998 (Rel. 36, Created)
DT 15-JUN-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ynfB precursor.
GN VNF OR B1583.
OS Escherichia coli.
CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC Enterobacteriaceae; Escherichia.
CC NCBI_TaxID=562;
CX [1]
RN SEQUENCE FROM N.A.
RA STRAIN-K12 / MG1655.
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000254; AAC74655.1; -.
DR PIR; A64914; A64914.
DR EcoGene; EG1840; ynfB.
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 113 HYPOTHETICAL PROTEIN VNF.
SQ SEQUENCE 113 AA; 12909 MW; 54CD49B384A7418 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 60;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy	11	LA11L	16
Db	11	LA11L	16

Search completed: January 21, 2004, 12:25:40  
Job time : 19 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: January 21, 2004, 12:23:52 ; Search time 35 Seconds  
(without alignments)  
1010.092 Million cell updates/sec

Title: US-09-941-314-2

Perfect score: 137  
1 MABPMQALQLLAITLTLMA.....VPAVPMFEQYKINKSCSSD 137

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: listing first 75 summaries

Database :

SPTREMBL.23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	137	100.0	138	4 O8WXU6	O8WXU6 homo sapien
2	75	54.7	103	4 O8WXU5	O8WXU5 homo sapien
3	8	5.8	245	2 O34115	O34115 lactococcus
4	8	5.8	601	5 O9VQR4	O9VQR4 drosophila
5	8	5.8	630	5 O8IA43	O8IA43 drosophila
6	8	5.8	707	16 O8YOP7	O8YOP7 anabaena sp
7	8	5.8	710	5 O9XX10	O9XX10 caenorhabdi
8	8	5.8	1713	5 O8MX19	O8MX19 caenorhabdi
9	8	5.8	1862	5 O95X73	O95X73 caenorhabdi
10	8	5.8	2836	5 O9U3B3	O9U3B3 caenorhabdi
11	8	5.8	2846	5 O9Y0A2	O9Y0A2 caenorhabdi
12	8	5.8	2847	5 O95X74	O95X74 caenorhabdi
13	8	5.8	2857	5 O9Y0A0	O9Y0A0 caenorhabdi
14	8	5.8	2892	5 O9Y0A1	O9Y0A1 caenorhabdi
15	8	5.8	2903	5 O95X75	O95X75 caenorhabdi
16	8	5.8	2933	5 O9GS15	O9GS15 caenorhabdi

17	7	5.1	24	8 O8SHB0	O8SHB0 rhampoleon
18	7	5.1	75	2 O51333	O51333 rhodobacter
19	7	5.1	89	16 O9J5K2	O9J5K2 chlamydia p
20	7	5.1	94	12 O8V6Y1	O8V6Y1 influenza a
21	7	5.1	94	12 O8V6Y2	O8V6Y2 influenza a
22	7	5.1	105	16 O929D1	O929D1 chlamydia p
23	7	5.1	108	12 O8OLW0	O8OLW0 influenza a
24	7	5.1	111	12 O91E55	O91E55 influenza v
25	7	5.1	112	12 O09684	O09684 influenza v
26	7	5.1	112	12 O09680	O09680 influenza v
27	7	5.1	112	12 O8JN18	O8JN18 influenza v
28	7	5.1	112	12 O09690	O09690 influenza a
29	7	5.1	112	12 O8OLR5	O8OLR5 influenza a
30	7	5.1	112	12 O09682	O09682 influenza v
31	7	5.1	112	12 O09692	O09692 influenza v
32	7	5.1	112	12 O09688	O09688 influenza v
33	7	5.1	112	12 O09686	O09686 influenza v
34	7	5.1	113	12 O8OLR3	O8OLR3 influenza a
35	7	5.1	114	12 O09698	O09698 influenza v
36	7	5.1	115	12 O8OLZ1	O8OLZ1 influenza a
37	7	5.1	115	12 O9QOG8	O9QOG8 influenza a
38	7	5.1	116	12 O8QLT8	O8QLT8 influenza a
39	7	5.1	116	12 O09703	O09703 influenza v
40	7	5.1	117	12 O9QOG6	O9QOG6 influenza a
41	7	5.1	118	12 O9QOB4	O9QOB4 influenza a
42	7	5.1	118	12 O09701	O09701 influenza v
43	7	5.1	118	12 O09694	O09694 influenza v
44	7	5.1	118	12 O9QOG4	O9QOG4 influenza a
45	7	5.1	119	12 O9QOF2	O9QOF2 influenza a
46	7	5.1	119	12 O9INP2	O9INP2 influenza a
47	7	5.1	119	12 O9QOP8	O9QOP8 influenza a
48	7	5.1	119	12 O9QOP4	O9QOP4 influenza a
49	7	5.1	119	12 O9QOG2	O9QOG2 influenza a
50	7	5.1	119	12 O9QOP0	O9QOP0 influenza a
51	7	5.1	119	12 O9QOB6	O9QOB6 influenza a
52	7	5.1	119	12 O9QOG0	O9QOG0 influenza a
53	7	5.1	119	12 O9QOB2	O9QOB2 influenza a
54	7	5.1	119	12 O9TP46	O9TP46 influenza a
55	7	5.1	119	12 O9TP44	O9TP44 influenza a
56	7	5.1	119	12 O9QOB8	O9QOB8 influenza a
57	7	5.1	119	12 O9QOP6	O9QOP6 influenza a
58	7	5.1	119	12 O9IBY7	O9IBY7 influenza a
59	7	5.1	120	12 O89726	O89726 influenza v
60	7	5.1	120	12 O67258	O67258 influenza v
61	7	5.1	120	12 O91R64	O91R64 influenza a
62	7	5.1	120	12 O09696	O09696 influenza v
63	7	5.1	121	12 O41656	O41656 influenza v
64	7	5.1	121	12 O8OPX6	O8OPX6 influenza a
65	7	5.1	121	12 O9WC22	O9WC22 influenza a
66	7	5.1	121	12 O99Y18	O99Y18 influenza a
67	7	5.1	121	12 O8Q758	O8Q758 influenza a
68	7	5.1	121	12 O57270	O57270 influenza v
69	7	5.1	121	12 O9PFW1	O9PFW1 influenza a
70	7	5.1	121	12 O41652	O41652 influenza a
71	7	5.1	121	12 O9Y274	O9Y274 influenza a
72	7	5.1	121	12 O8Q763	O8Q763 influenza a
73	7	5.1	121	12 O9YJF2	O9YJF2 influenza a
74	7	5.1	121	12 O91U30	O91U30 influenza a
75	7	5.1	121	12 O82806	O82806 influenza v

## ALIGNMENTS

RESULT 1	O8WXU6	PRELIMINARY;	PRT;	138 AA.
AC	O8WXU6;			
DT	01-MAR-2002 (TREMBlrel. 20, Created)			
DT	01-MAR-2002 (TREMBlrel. 20, Last sequence update)			
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)			
DE	SC13.			
OS	Homo sapiens (Human).			

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Hamil K.G., Liu Q., Zhang Y.-L., French F.S., Hall S.H.;
RT "SC13: A novel epididymal specific member of the cystatin family.";
RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF35480; AAL71991.1; -
DR InterPro; IPR00010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
DR SMART; SM00043; CY; 1.
SQ SEQUENCE 138 AA; 16506 MW; E49440ACA3585C64 CRC64;

Query Match 100.0%; Score 137; DB 4; Length 138;
Best Local Similarity 100.0%; Pred. No. 9.7e-141;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEPWALQILALITLMAIPYQARKTFLSVHVAENVAYADSLQWITDQYNKESDD 60
DB 2 MAEPWALQILALITLMAIPYQARKTFLSVHVAENVAYADSLQWITDQYNKESDD 61
QY 61 KYHFRIFRVLKVRQVTDHLEHVLNEMQWTTCKPFTTNCVPOREILHKVNCFFSVFA 120
DB 62 KYHFRIFRVLKVRQVTDHLEHVLNEMQWTTCKPFTTNCVPOREILHKVNCFFSVFA 121
QY 121 VPMFQYKILNKSSD 137
DB 122 VPMFQYKILNKSSD 138

RESULT 2
08WKUS PRELIMINARY; PRT; 103 AA.
AC 08WKUS;
DT 01-MAR-2002 (TREMBlrel. 20; Created)
DT 01-MAR-2002 (TREMBlrel. 20; Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21; Last annotation update)
DE SC13delta.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Hamil K.G., Liu Q., Zhang Y.-L., French F.S., Hall S.H.;
RT "SC13: A novel epididymal specific member of the cystatin family.";
RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF35481; AAL71992.1; -
DR InterPro; IPR00010; Cystatin.
DR Pfam; PF00031; cystatin; 1.
SQ SEQUENCE 103 AA; 12285 MW; 05DD92C47387B022 CRC64;

Query Match 54.7%; Score 75; DB 4; Length 103;
Best Local Similarity 100.0%; Pred. No. 2e-73;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEPWALQILALITLMAIPYQARKTFLSVHVAENVAYADSLQWITDQYNKESDD 60
DB 2 MAEPWALQILALITLMAIPYQARKTFLSVHVAENVAYADSLQWITDQYNKESDD 61
QY 61 KYHFRIFRVLKVRQ 75
DB 62 KYHFRIFRVLKVRQ 76

RESULT 3
034115 PRELIMINARY; PRT; 245 AA.
AC 034115;
DT 01-JAN-1998 (TREMBlrel. 05; Created)
DT 01-JAN-1998 (TREMBlrel. 05; Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19; Last annotation update)

```

```

DE lctg.
GN lctg.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Plasmid pBS2.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C2102;
RC MEDLINE=98027362; PubMed=9361411;
RA Rince A., Dufour A., Uguen P., Le Pennec J.P., Haras D.;
RT "Characterization of the lactacin 481 operon: the Lactococcus lactis genes lctf, lctg, and lctg encode a putative ABC transporter involved in bacteriocin immunity.";
RL Appl. Environ. Microbiol. 63:4252-4260(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=C2102;
RC MEDLINE=20416241; PubMed=10960114;
RA Uguen P., Le Pennec J.P., Dufour A.;
RT "Antibiotic biosynthesis: interactions between prelactacin 481 and its putative modification enzyme, lctm.";
RL J. Bacteriol. 182:5262-5266(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=C2102;
RC MEDLINE=20444212; PubMed=10986268;
RA Dufour A., Rince A., Uguen P., Le Pennec J.P.;
RT "IS1675, a Novel Lactococcal Insertion Element, Forms a Transposon-Like Structure Including the Lactacin 481 Antibiotic Operon.";
RL J. Bacteriol. 182:5600-5605(2000).
DR EMBL; U91581; AAC72255.1; -
KW Plasmid.
SQ SEQUENCE 245 AA; 28420 MW; 496FA9A23D874F1E CRC64;

Query Match 5.8%; Score 8; DB 2; Length 245;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ALQILAI 14
DB 157 ALQILAI 164

RESULT 4
09VQRA PRELIMINARY; PRT; 601 AA.
AC 09VQRA;
DT 01-MAY-2000 (TREMBlrel. 13; Created)
DT 01-OCT-2002 (TREMBlrel. 22; Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23; Last annotation update)
DE CG31776 protein.
OS Drosophila melanogaster (fruit fly).
GN CG31956 OR CG8845.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Berkley;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoekstra R.A., Galie R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutcliffe R., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley B.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokslein P., Brotter P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

```

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Dugas A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.B., Downes M., Dyan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Engelilista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fester C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaser K.,  
RA Glieder A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris N.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalish P., Kapten G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kammel B.E., Kodire C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Laeko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mlehnina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paciel J.M.,  
RA Palazzolo M., Piltman G.S., Pan S., Pollard J., Puti V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupeki M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svaythae R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach U.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Centker S.E., Adams M.D., Krommler B., Wan K.H., Holt R.A.,  
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Buesam D.A.,  
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
RA Dodson K., Dorett V., Doup L.B., Doyle C., Drenek D., Farfan D.,  
RA Ferreira S., Frise E., Galle R.P., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Idegaw C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Paciel J., Parasag V., Park S., Patel S., Pfeiffer B.,  
RA Phouenavong S., Piltman G.S., Puti V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svaythae R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RT "Sequencing of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Betman B., Carlson J.W., Celniker S.E.,  
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Krommler B., Marshall B., Milburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith B., Shu S., Smutnick F., Whitfield B.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungell C.J., Lewis S.E.;  
RT "Annotation of *Drosophila melanogaster* genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
RA EMBL, AB003579; AAF51101.2; -  
DR FlyBase; FBN0051776; CG31776.  
DR FlyBase; FBN0051956; CG31956.  
DR InterPro; IPR001173; Glyco\_tran\_2.  
DR InterPro; IPR000772; Ricin\_B lectin.  
DR Pfam; PF00535; Glycos\_transf\_2; 1.  
DR Pfam; PF00652; Ricin\_B lectin; 2.  
DR SMART; SM00458; RICIN\_1.  
DR PROSITE; PSS0231; RICIN\_B\_LECTIN; 1.  
DR PROSITE; PSS0231; RICIN\_B\_LECTIN; 1834526G43CEFF81 CRC64;  
QO SEQUENCE 601 AA; 69404 MW; 1834526G43CEFF81 CRC64;

Query Match	5.8%;	Score 8;	DB 5;	Length 60;
Best Local Similarity	100.0%;	Pred. No. 6.5;		
Matches	8;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	10 LLLALLLT 17			
DB	11 LLLALLLT 18			
RESULT 5				
OBIA43	PRELIMINARY;	PRT;	630 AA.	
AC	Q8IA43.			
DT	01-MAR-2003 (TrEMBLrel. 23, Created)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)			
DE	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)			
DE	N-acetylglactosaminyltransferase.			
CN	C68845.1			
OS	Drosophila melanogaster (fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
CC	NCBI_Taxid=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Seewientek T.J., Clausen H.;			
RT	"A Genome Survey of UDP-N-acetylglactosamine: Polypeptide N-			
RT	acetylglactosaminyltransferases of Drosophila."			
RT	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.			
KW	EMBL; AF324751; AAN75750.1; -.			
DR	Transferase			
SO	SEQUENCE 630 AA; 72795 MW; 7B2748165A9293CA CRC64;			
Query Match				
Best Local Similarity	5.8%;	Score 8;	DB 5;	Length 630;
Matches	8;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	10 LLLALLLT 17			
DB	11 LLLALLLT 18			
RESULT 6				
OBXP7	PRELIMINARY;	PRT;	707 AA.	
ID	Q8Y0P7			
AC	Q8Y0P7.			
DT	01-MAR-2002 (TrEMBLrel. 20, Created)			
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)			
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)			
DE	Serine/threonine kinase.			
CN	ALJ3773.			
OS	Anabaena sp. (strain PCC 7120).			
OC	Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.			
CC	NCBI_TaxID=103690;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE=21595285; PubMed=11759840;			
RA	Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Saeamoto S.;			
RA	Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Saeamoto S.;			
RA	Kishida Y., Kohara M., Matsuno M., Matsuno A., Muraki A.;			
RA	Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.;			
RA	Yasuda M., Tabata S.;			
RT	"Complete genomic sequence of the filamentous nitrogen-fixing			
RT	cyanobacterium Anabaena sp. strain PCC 7120."			
RT	DNA Ref. 8:205-213 (2001).			
DR	EMBL; AF003594; BAB75472.1; -.			
DR	InterPro; IPR000719; Prot_Kinase.			
DR	InterPro; IPR001440; TPR.			
DR	Pfam; PF00069; Pkinase; 1.			
DR	Pfam; PF00515; TPR; 10.			
DR	SMART; SM0028; TPR; 10.			
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP_1.			

DR PROSITE; PS50011; PROTEIN KINASE\_DOM; 1.  
KM Kinase; Complete proteome-  
SQ SEQUENCE 707 AA; 80375 MW; 2D1B540ADE5551C CRC64;

Query Match 5.8%; Score 8; DB 16; Length 707;  
Best Local Similarity 100.0%; Pred. No. 7.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 LLLALLLT 17  
Db 303 LLLALLLT 310

## RESULT 7

Q9XX10 PRELIMINARY; PRT; 710 AA.

AC Q9XX10; 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DE Y39A1A.22 protein.

GN Y39A1A.22.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;

RN [1]  
RF SEQUENCE FROM N.A.

RA Wall M.;  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

RN [2]  
RP SEQUENCE FROM N.A.

RX MEDLINE=99069613; PubMed=9851916;

RA none;  
RT "Genome sequence of the nematode C. elegans: A platform for

RT Investigating biology.";

RL Science 282:2012-2018(1998).

DR EMBL; AL031633; CAA21031.1; -  
DR WormPep; Y39A1A.22; CE19142.

DR InterPro; IPR004342; EKS\_cterm.  
DR InterPro; IPR004331; SPX.

DR Pfam; PF03124; EKS; 1.  
DR Pfam; PF03105; SPX; 1.

SQ SEQUENCE 710 AA; 82375 MW; 772DC887594E3258 CRC64;

Query Match 5.8%; Score 8; DB 5; Length 710;  
Best Local Similarity 100.0%; Pred. No. 7.5;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 LLLALLLT 17  
Db 238 LLLALLLT 245

## RESULT 8

Q8MX19 PRELIMINARY; PRT; 1713 AA.

AC Q8MX19; 01-OCT-2002 (TREMBLrel. 22, Created)  
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE ITR-1 protein (corresponding sequence F33D4.2g).

GN F33D4.2 OR ITR-1.  
OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]  
RF SEQUENCE FROM N.A.

RA STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA Waterston R.;  
RT "Genome sequence of the nematode C. elegans: a platform for

RT Investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).

RN [2]  
RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;  
RA Stellyes L.; Johnson D.;

RT "The sequence of C. elegans cosmid F33D4.";  
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.

RN [3]  
RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;  
RA Waterston R.;

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF036702; AAM98028.1; -  
DR WormPep; F33D4.2g; CE31690.

DR InterPro; IPR000699; Ca-rel\_channel.  
DR InterPro; IPR000493; Insp3\_receptor.

DR Pfam; PF02815; MIR; 2.  
DR Pfam; PF01365; RYDR\_TPR; 2.

DR PRINTS; PR00779; INSP3RECEPT.  
DR SMART; SM00472; MIR; 3.

SQ SEQUENCE 1713 AA; 196075 MW; 5FEDC945C48A2A2C CRC64;

Query Match 5.8%; Score 8; DB 5; Length 1713;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLLALLLT 16  
Db 1093 QLLALLLT 1100

## RESULT 9

Q95X73 PRELIMINARY; PRT; 1862 AA.

AC Q95X73; 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE Hypothetical 213.0 kDa protein.

GN F33D4.2.

OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;

RN [1]  
RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;  
RA STRAIN=Bristol N2;

RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF036702; AAK68363.1; -  
DR WormPep; F33D4.2c; CE28015.

DR InterPro; IPR000699; Ca-rel\_channel.  
DR InterPro; IPR000493; Insp3\_receptor.

DR Pfam; PF02815; MIR; 2.  
DR Pfam; PF01365; RYDR\_TPR; 2.

SQ SEQUENCE 1713 AA; 196075 MW; 5FEDC945C48A2A2C CRC64;

Query Match 5.8%; Score 8; DB 5; Length 1713;  
Best Local Similarity 100.0%; Pred. No. 16;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLLALLLT 16  
Db 1093 QLLALLLT 1100

DR PRINTS; PR00779; INSP3RECEPTR.  
 DR SMART; SM00472; MIR; 3.  
 KW Hypothetical protein.  
 SQ SEQUENCE 1862 AA; 213026 MW; E6B07F3817DDB03 CRC64;  
 Query Match 5.8%; Score 8; DB 5; Length 1862;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLLAILL 16  
 Db 1093 QLLAILL 1100

RESULT 10  
 Q9U3B3 PRELIMINARY; PRT; 2836 AA.  
 ID Q9U3B3;  
 AC Q9U3B3;  
 DT 01-MAY-2000 (TEMBUREL. 13, Created)  
 DT 01-MAY-2000 (TEMBUREL. 13, Last sequence update)  
 DT 01-OCT-2002 (TEMBUREL. 22, Last annotation update)  
 DE Inositol 1,4,5-trisphosphate receptor.  
 GN ITR-1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=20079510; PubMed=10610772;  
 RA Baylis H.A., Furutachi T., Yoshikawa F., Mikoshiba K., Sattelle D.B.,  
 RT "Inositol 1,4,5-trisphosphate Receptors are strongly expressed in the  
 RT nervous system, pharynx, intestine, gonad and excretory cell of  
 RT Caenorhabditis elegans and are encoded by a single gene (itr-1).";  
 RL J. Mol. Biol. 294:467-476(1999).  
 DR EMBL; AJ243181; CAB45862.1; -;  
 DR InterPro: IPR000699; Ca-rel\_channel.  
 DR InterPro: IPR001682; Ca/Na\_pore.  
 DR InterPro: IPR000493; Insp3\_receptor.  
 DR InterPro: IPR005821; Ion\_trans.  
 DR InterPro: IPR003608; MIR.  
 DR Pfam; PF00520; Ion\_trans; 1.  
 DR Pfam; PF02815; MIR; 2.  
 DR Pfam; PF01365; RYDR\_ITPR; 2.  
 DR PRINTS; PR00779; INSP3RECEPTR.  
 DR SMART; SM00472; MIR; 3.  
 KW Ionic channel; Receptor; Transmembrane.  
 SQ SEQUENCE 2836 AA; 324561 MW; 3A2385BC8AC870F9 CRC64;  
 Query Match 5.8%; Score 8; DB 5; Length 2836;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLLAILL 16  
 Db 1083 QLLAILL 1090

RESULT 11  
 Q9Y0A2 PRELIMINARY; PRT; 2846 AA.  
 ID Q9Y0A2;  
 AC Q9Y0A2;  
 DT 01-NOV-1999 (TEMBUREL. 12, Created)  
 DT 01-NOV-1999 (TEMBUREL. 12, Last sequence update)  
 DT 01-MAR-2003 (TEMBUREL. 23, Last annotation update)  
 DE Inositol 1,4,5-trisphosphate receptor (Hypothetical 325.9 kDa  
 DE protein).  
 GN ITR-1 OR F33D4.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Baylis H.A., Furutachi T., Yoshikawa F., Mikoshiba K., Sattelle D.B.,  
 RT "Inositol 1,4,5-trisphosphate Receptors are strongly expressed in the  
 RT nervous system, pharynx, intestine and excretory cell of  
 RT Caenorhabditis elegans and are encoded by a single gene (itr-1).";  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99428139; PubMed=10499793;  
 RA Dal Santo P., Logan M.A., Chisholm A.D., Jorgensen B.M.,  
 RT "The inositol trisphosphate receptor regulates a 50-second behavioral  
 RT rhythm in C. elegans.";  
 RL Cell 98:757-767(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99065613; PubMed=9851916;  
 RA "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Stellyes L., Johnson D.,  
 RT "The sequence of C. elegans cosmid F33D4.";  
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.,  
 RT "Direct Submission.";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ243179; CAB45860.1; -;  
 DR EMBL; AF168688; AAF05302.1; -;  
 DR EMBL; AF036702; AAK68366.1; -;  
 DR WormPep; F33D4.2; CR28018.  
 DR InterPro: IPR000699; Ca-rel\_channel.  
 DR InterPro: IPR001682; Ca/Na\_pore.  
 DR InterPro: IPR000493; Insp3\_receptor.  
 DR InterPro: IPR005821; Ion\_trans.  
 DR InterPro: IPR003608; MIR.  
 DR Pfam; PF00520; Ion\_trans; 1.  
 DR Pfam; PF02815; MIR; 2.  
 DR Pfam; PF01365; RYDR\_ITPR; 2.  
 DR PRINTS; PR00779; INSP3RECEPTR.  
 DR SMART; SM00472; MIR; 3.  
 KW Hypothetical protein; Ionic channel; Receptor; Transmembrane.  
 SQ SEQUENCE 2846 AA; 325911 MW; 0186413338AB3BA2 CRC64;  
 Query Match 5.8%; Score 8; DB 5; Length 2846;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QLLAILL 16  
 Db 1093 QLLAILL 1100

RESULT 12  
 Q95X74 PRELIMINARY; PRT; 2847 AA.  
 ID Q95X74;  
 AC Q95X74;  
 DT 01-DEC-2001 (TEMBUREL. 19, Created)  
 DT 01-DEC-2001 (TEMBUREL. 19, Last sequence update)  
 DT 01-MAR-2003 (TEMBUREL. 23, Last annotation update)  
 DE Hypothetical 325.8 kDa protein.  
 GN F33D4.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.

```

RT RT "genome sequence of the nematode C. elegans: a platform for
RT RT investigating biology. The C. elegans Sequencing Consortium.";
RL RL Science 282:2012-2018(1998).
RN RN [3]
RP RP SEQUENCE FROM N.A.
RC RC STRAIN=Bristol N2;
RA RA Stelleyes L., Johnson D.;
RT RT "the sequence of C. elegans cosmid F3Jd.";
RL RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
RN RN [4]
RP RP SEQUENCE FROM N.A.
RC RC STRAIN=Bristol N2;
RA RA Waterston R.;
RT RT "Direct Submission.";
RL RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN RN [5]
RP RP EMBL; AJ243182; CAB45863.1; -.
DR DR EMBL; AF036702; AAK68365.1; -.
DR DR WormPeP; F3Jd.2e; CE28017.
DR DR InterPro; IPR000699; Ca-rel channel.
DR DR InterPro; IPR000493; Insp3_receptor.
DR DR InterPro; IPR005821; Ion_trans.
DR DR Pfam; PF00520; Ion_trans; 1.
DR DR Pfam; PF02815; MIR_2.
DR DR Pfam; PF0365; RYDR_ITPR; 2.
DR DR PRINTS; PR00779; INSP3RECEPTR.
DR DR SMART; SMO0472; MIR; 3.
DR DR Hypothetical protein; Ionic channel; Receptor; Transmembrane.
SQ SQ SEQUENCE 2857 AA; 327195 MW; 8B991F12B8C94E2F CRC64;

Query Match 5.8%; Score 8; DB 5; Length 2857;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 QLTATL 16
DB 1102 QLTATL 1109

RESULT 14
O9Y0A1 PRELIMINARY; PRT; 2892 AA.
AC O9Y0A1;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, last annotation update)
DE Intercol 1,4,5-trisphosphate receptor.
GN ITR-1.
OC Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN RN [1]
RP RP SEQUENCE FROM N.A.
RC RC STRAIN=Bristol N2;
RX RX MEDLINE=20079510; PubMed=10610772;
RA Rayliss H.A., Fujiuchi T., Yoshikawa F., Mikoshiba K., Sattelle D.B.;
RT RT "Inositol 1,4,5-trisphosphate Receptors are strongly expressed in the
RT RT nervous system, pharynx, intestine, gonad and excretory cell of
RT RT Caenorhabditis elegans and are encoded by a single gene (itr-1).";
RL RL J. Mol. Biol. 294:467-476(1999).
DR DR EMBL; AJ243180; CAB45861.1; -.
DR DR InterPro; IPR000699; Ca-rel channel.
DR DR InterPro; IPR001682; Ca/Na_pore.
DR DR InterPro; IPR000493; Insp3_receptor.
DR DR InterPro; IPR003608; MIR.
DR DR Pfam; PF00520; Ion_trans; 1.
DR DR Pfam; PF02815; MIR_2.
DR DR Pfam; PF0365; RYDR_ITPR; 2.
DR DR PRINTS; PR00779; INSP3RECEPTR.
DR DR SMART; SMO0472; MIR; 3.

```

KW Ionic channel; Receptor; Transmembrane.  
SQ SEQUENCE 2892 AA; 330832 MW; B1C598F8554B83CF CRC64;

Query Match 5.8%; Score 8; DB 5; Length 2892;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QULLATL 16  
DB 1139 QULLATL 1146

## RESULT 15

Q95X75 PRELIMINARY; PRT; 2903 AA.  
AC Q95X75;  
DT 01-DEC-2001 (T-EMBLrel. 19, Created)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
DE Hypothetical 332.1 kDa protein.  
GN F33D4.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology. The C. elegans Sequencing Consortium.";  
RT Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Bristol N2;  
RA Stellyes L., Johnson D.;  
RT "The sequence of C. elegans cosmid F33D4.";  
RT Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Bristol N2;  
RA Waterston R.;  
RT "Direct Submission.";  
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
RN EMBL; AF036702; AAK68362.1; -;  
DR WormBep: F33D4.2a; CE28014.  
DR InterPro: IPR000669; Ca-rel\_channel.  
DR InterPro: IPR001682; Ca/Na\_pore.  
DR InterPro: IPR000493; Insp3\_receptor.  
DR InterPro: IPR005821; Ion\_trans.  
DR InterPro: IPR003608; MIR.  
DR Pfam: PF00520; Ion\_trans; 1.  
DR Pfam: PF02815; MIR\_2.  
DR Pfam: PF01365; RYDR\_TTPR; 2.  
DR PRINTS; PR00779; INSP3RECEPTR.  
DR SMART; SM00472; MIR; 3.  
KM Hypothetical protein; Ionic channel; Transmembrane.  
SQ SEQUENCE 2903 AA; 332089 MW; 22B738A5D0896872 CRC64;

Query Match 5.8%; Score 8; DB 5; Length 2903;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QULLATL 16  
DB 1139 QULLATL 1146

RESULT 16  
Q9GS15 PRELIMINARY; PRT; 2933 AA.  
AC Q9GS15;

DT 01-MAR-2001 (T-EMBLrel. 16, Created)  
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)  
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)  
DE Putative inositol 1,4,5-trisphosphate receptor.  
GN ITR-1.  
OS Caenorhabditis briggsae.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6238;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AP16;  
RA Howard B.A.;  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-AP16;  
RX MEDLINE=21136503; PubMed=11237590;  
RA Gower N.J.D., Temple G.R., Schein J.B., Marra M., Walker D.S.,  
RA Baylis H.A.;  
RT "Dissection of the promoter region of the inositol 1,4,5-trisphosphate  
RT receptor gene, Itr-1, in C. elegans: A molecular basis for cell-  
RT specific expression of IP3R isoforms.";  
RL J. Mol. Biol. 306:145-157(2001).  
RN EMBL; AJ293919; CAC10528.1; -;  
DR InterPro: IPR000669; Ca-rel\_channel.  
DR InterPro: IPR001682; Ca/Na\_pore.  
DR InterPro: IPR000493; Insp3\_receptor.  
DR InterPro: IPR005821; Ion\_trans.  
DR InterPro: IPR003608; MIR.  
DR Pfam: PF00520; Ion\_trans; 1.  
DR Pfam: PF02815; MIR\_2.  
DR Pfam: PF01365; RYDR\_TTPR; 2.  
DR PRINTS; PR00779; INSP3RECEPTR.  
DR SMART; SM00472; MIR; 3.  
KM Ionic channel; Receptor; Transmembrane.  
SQ SEQUENCE 2933 AA; 335223 MW; 0B307A206345C2FB CRC64;

Query Match 5.8%; Score 8; DB 5; Length 2933;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QULLATL 16  
DB 1153 QULLATL 1160

RESULT 17  
Q8SHB0 PRELIMINARY; PRT; 24 AA.  
AC Q8SHB0;  
DT 01-JUN-2002 (T-EMBLrel. 21, Created)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
DE NADH dehydrogenase subunit 1 (Fragment).  
GN NDI.  
OS Rhaphiolepis spectrum.  
OC Mitochondrion.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Lepidosteiformes; Squamata; Iguania; Acrodonta; Chamaeleonidae;  
OC Rhaphiolepis.  
OX NCBI\_TaxID=179929;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Townsend T.M., Larson A.L.;  
RT "Molecular Phylogenetics and Mitochondrial Genomic Evolution in the  
RT Chamaeleonidae (Reptilia, Squamata).";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF448772; AAL90596.1; -;  
KM Mitochondrion.  
FT NON TER 1  
SQ SEQUENCE 24 AA; 2666 MW; C95A1FA4AB3791CB CRC64;

Query Match 5.1%; Score 7; DB 8; Length 24;  
 Best Local Similarity 100.0%; Pred. No. 5.1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 LTMALP 22  
 |||||  
 DB 18 LTMALP 24

## RESULT 18

Q53133 PRELIMINARY; PRT; 75 AA.  
 AC Q53133;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE ORFX protein.  
 GN ORFX.  
 OS Rhodospirillum rubrum (Rhodospirillum rubrum).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;  
 OC Rhodospirillaceae; Rhodospirillum.  
 OX NCBI\_TaxID=1063;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WS8-N;  
 RA MEDLINE=96079285; PubMed=7494484;  
 RA Ward M.J., Bell A.W., Hamblin P.A., Packer H.L., Armitage J.P.;  
 RT "Identification of a chemotaxis operon with two cheY genes in  
 RT Rhodospirillum rubrum."  
 RL Mol. Microbiol. 17:357-366(1995).  
 DR EMBL; X80027; CAA56328.1; -  
 SQ SEQUENCE 75 AA; 8189 MW; 748F5047EA4718CF CRC64;

Query Match 5.1%; Score 7; DB 2; Length 75;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ALQLLA 13  
 |||||  
 DB 26 ALQLLA 32

## RESULT 19

Q9J5K2 PRELIMINARY; PRT; 89 AA.  
 AC Q9J5K2;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE Hypothetical protein CPJ0050.  
 GN CPJ0050.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J138;  
 RA MEDLINE=20330349; PubMed=10871362;  
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
 RT from Japan and CML029 from USA."  
 RL Nucleic Acids Res. 28:2311-2314(2000).  
 DR EMBL; AF002545; BAA98261.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 89 AA; 9832 MW; 7A16851430490254 CRC64;

Query Match 5.1%; Score 7; DB 16; Length 89;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLAAIL 16  
 |||||

DB 60 LLAAIL 66

## RESULT 20

Q8V6Y1 PRELIMINARY; PRT; 94 AA.  
 AC Q8V6Y1;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Non-structural protein (Fragment).  
 GN NS.  
 OS Influenza A virus (A/Swine/Italy/25823/94(H3N2)).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza A viruses; Influenzavirus A.  
 OX NCBI\_TaxID=147095;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A/Swine/Italy/25823/94(H3N2);  
 RA Campitelli I., Donatelli I., Fabiani C., Puzelli S., Fioretti A.,  
 RA Foni E., De Marco A., Kraus S., Webster R.G.;  
 RT "H3N2 influenza viruses from domestic chickens in Italy: An increasing  
 RT role for chickens in the ecology of influenza?"  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF213908; AAU59056.1; -  
 DR InterPro; IPR000968; Flu\_NS2.  
 DR Pfam; PF00601; Flu\_NS2; I.  
 FT NON TER 1  
 SQ SEQUENCE 94 AA; 11347 MW; 2A8B2DEDDEDDP232C CRC64;

Query Match 5.1%; Score 7; DB 12; Length 94;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLL 12  
 |||||  
 DB 74 QALQLL 80

## RESULT 21

Q8V6Y2 PRELIMINARY; PRT; 94 AA.  
 AC Q8V6Y2;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Non-structural protein (Fragment).  
 GN NS.  
 OS Influenza A virus (A/Mallard/Italy/24/95(H1N1)).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza A viruses; Influenzavirus A.  
 OX NCBI\_TaxID=147097;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A/Mallard/Italy/24/95(H1N1);  
 RA Campitelli I., Donatelli I., Fabiani C., Puzelli S., Fioretti A.,  
 RA Foni E., De Marco A., Kraus S., Webster R.G.;  
 RT "H3N2 influenza viruses from domestic chickens in Italy: An increasing  
 RT role for chickens in the ecology of influenza?"  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF213907; AAU59055.1; -  
 DR InterPro; IPR000968; Flu\_NS2.  
 DR Pfam; PF00601; Flu\_NS2; I.  
 FT NON TER 1  
 SQ SEQUENCE 94 AA; 11299 MW; 4E7B3ABC50E36E8D CRC64;

Query Match 5.1%; Score 7; DB 12; Length 94;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLL 12  
 |||||  
 DB 74 QALQLL 80

RESULT 22

Q929D1 PRELIMINARY; PRT; 105 AA.

AC Q929D1;

DT 01-MAY-1999 (TrEMBLrel. 10, Created)

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)

DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

GN Hypothetical protein Cpn0050.

OS Chlamydia pneumoniae (Chlamydia pneumoniae).

OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI\_TaxID=83558;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CWL029;

RX MEDLINE=99206606; PubMed=10192388;

RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W., Olinger L., Grimwood J., Davis R.W., Stephens R.S., "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";

RT Nat. Genet. 21:385-389(1999).

RL EMBL; AF001590; AAD18203.1; -.

DR InterPro; IPR000968; Flu\_NS2.

DR Pfam; PF00601; Flu\_NS2; I.

FT NON\_TER 1

FT NON\_TER 108

FT SEQUENCE 108 AA; 12822 MW; A50D6D5679C24677 CRC64;

Query Match 5.1%; Score 7; DB 16; Length 105;

Best Local Similarity 100.0%; Pred. No. 18;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLALL 16

DB 60 LLLALL 66

RESULT 23

Q8QLW0 PRELIMINARY; PRT; 108 AA.

AC Q8QLW0;

DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

GN NS2.

OS Influenza A virus (A/swine/Cotes d'Armor/604/99 (H1N2)).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza A viruses; Influenzavirus A.

OX NCBI\_TaxID=158307;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A/swine/Cotes d'Armor/604/99 (H1N2);

RX MEDLINE=21904432; PubMed=11907321;

RA Marzin S., Gregory V., Cameron K., Bennett M., Valette M., Aymer M., Font E., Barigazzi G., Lin Y., Hay A., "Antigenic and genetic diversity among swine influenza A H1N1 and H1N2 viruses in Europe.";

RT J. Gen. Virol. 83:735-745(2002).

RL EMBL; AJ316062; CAC87415.1; -.

DR InterPro; IPR000968; Flu\_NS2.

DR Pfam; PF00601; Flu\_NS2; I.

FT NON\_TER 1

FT NON\_TER 108

FT SEQUENCE 108 AA; 12822 MW; A50D6D5679C24677 CRC64;

Query Match 5.1%; Score 7; DB 12; Length 108;

Best Local Similarity 100.0%; Pred. No. 18;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQALL 12

DB 96 QALQALL 102

RESULT 24

Q91E55 PRELIMINARY; PRT; 111 AA.

AC Q91E55;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

GN NS2.

OS Influenza virus.

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Unclassified Orthomyxoviridae.

OX NCBI\_TaxID=11309;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A/swine/Italy/636/87;

RX MEDLINE=21262224; PubMed=11369884;

RA Gregory V., Lim W., Cameron K., Bennett M., Klimov A., Hall H., Cox N., Hay A.J., Lin Y., "Infection of a child in Hong Kong by an influenza A H3N2 virus closely related to viruses circulating in European pigs.";

RT J. Gen. Virol. 82:1397-1406(2001).

RL EMBL; AJ293940; CAC40060.1; -.

DR InterPro; IPR000968; Flu\_NS2.

DR Pfam; PF00601; Flu\_NS2; I.

FT NON\_TER 1

FT NON\_TER 111

FT SEQUENCE 111 AA; 13252 MW; 9C3EA9C0AP91P10D CRC64;

Query Match 5.1%; Score 7; DB 12; Length 111;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQALL 12

DB 97 QALQALL 103

RESULT 25

Q09684 PRELIMINARY; PRT; 112 AA.

AC Q09684;

DT 01-JUL-1997 (TrEMBLrel. 04, Created)

DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)

DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

GN NS.

OS Influenzavirus A.

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza A viruses.

OX NCBI\_TaxID=197911;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A/swine/Iowa/17672/88;

RX MEDLINE=97048087; PubMed=8892928;

RA Guan Y., Shortridge K.F., Krause S., Li P.H., Kawaka Y., Webster R.G., "Emergence of avian H1N1 influenza viruses in pigs in China.";

RT J. Virol. 70:8041-8046(1996).

RL EMBL; U94985; AAB50998.1; -.

DR InterPro; IPR000968; Flu\_NS2.

DR Pfam; PF00601; Flu\_NS2; I.

FT NON\_TER 1

FT NON\_TER 112

FT SEQUENCE 112 AA; 13333 MW; D9CCED9F9059275B CRC64;

Query Match 5.1%; Score 7; DB 12; Length 112;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQALL 12

DB 98 QALQALL 104

## RESULT 26

```

009680 ID 009680 PRELIMINARY; PRT; 112 AA.
AC 009680;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Nonstructural protein NS2 (Fragment).
GN NS.
OS Influenzavirus A.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
CX NCBI_TaxID=197911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Ruddy Turnstone/New Jersey/47/85;
RX MEDLINE=97048087; PubMed=8892928;
RA Guan Y., Shortridge K.F., Krause S., Li P.H., Kawoka Y.,
  Webster R.G.;
RT "Emergence of avian H1N1 influenza viruses in pigs in China.";
RL J. Virol. 70:8041-8046(1996).
DR EMBL; U49483; AAB50994.1; -
DR InterPro; IPR000968; Flu_NS2.
DR Pfam; PF00601; Flu_NS2; 1.
FT NON_TER 1
FT SEQUENCE 112 AA; 13269 MW; D6628F0A595757F5 CRC64;
SQ

```

```

Query Match 5.1%; Score 7; DB 12; Length 112;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 6 QALQLL 12
DB 98 QALQLL 104

```

## RESULT 27

```

08JN18 ID 08JN18 PRELIMINARY; PRT; 112 AA.
AC 08JN18;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Nonstructural protein 2 (Fragment).
GN NS2.
OS Influenza A virus (A/Guangzhou/333/99 (H9N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
CX NCBI_TaxID=172856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Guangzhou/333/99 (H9N2);
RA Guo Y., Dong J., Wu K., Guo J., Wang M., Zhang L., Chen J.;
RT "Genetic characteristics of Influenza H9N2 subtype viruses isolated
  from human.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY033027; AAL32489.1; -
DR InterPro; IPR000968; Flu_NS2.
DR Pfam; PF00601; Flu_NS2; 1.
FT NON_TER 1
FT NON_TER 112
FT SEQUENCE 112 AA; 13274 MW; 331ADF26C61DB4A7 CRC64;
SQ

```

```

Query Match 5.1%; Score 7; DB 12; Length 112;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 6 QALQLL 12
DB 97 QALQLL 103

```

## RESULT 28

```

009690 ID 009690 PRELIMINARY; PRT; 112 AA.
AC 009690;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Nonstructural protein NS2 (Fragment).
GN NS.
OS Influenzavirus A.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
CX NCBI_TaxID=197911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Laughing Gull/Delaware/2838/87;
RX MEDLINE=97048087; PubMed=8892928;
RA Guan Y., Shortridge K.F., Krause S., Li P.H., Kawoka Y.,
  Webster R.G.;
RT "Emergence of avian H1N1 influenza viruses in pigs in China.";
RL J. Virol. 70:8041-8046(1996).
DR EMBL; U49488; AAB51004.1; -
DR InterPro; IPR000968; Flu_NS2.
DR Pfam; PF00601; Flu_NS2; 1.
FT NON_TER 1
FT NON_TER 112
FT SEQUENCE 112 AA; 13281 MW; 3BF6D5287770F94E CRC64;
SQ

```

```

Query Match 5.1%; Score 7; DB 12; Length 112;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 6 QALQLL 12
DB 98 QALQLL 104

```

## RESULT 29

```

08QLR5 ID 08QLR5 PRELIMINARY; PRT; 112 AA.
AC 08QLR5;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Non structural protein 2 (Fragment).
GN NS2.
OS Influenza A virus (A/swine/Cotes d'Armor/799/00 (H1N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
CX NCBI_TaxID=168284;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/swine/Cotes d'Armor/799/00 (H1N2);
RX MEDLINE=11904432; PubMed=11907321;
RA Marozin S., Gregory V., Cameron K., Bennett M., Valette M.,
  Font B., Barigazzi G., Lin Y., Hay A.;
RT "Antigenic and genetic diversity among swine influenza A H1N1 and H1N2
  viruses in Europe.";
RL J. Gen. Virol. 83:735-745(2002).
DR EMBL; AJ344040; CAC86658.1; -
DR InterPro; IPR00256; Flu_NS1.
DR InterPro; IPR000968; Flu_NS2.
DR Pfam; PF00600; Flu_NS1; 1.
DR Pfam; PF00601; Flu_NS2; 1.
FT NON_TER 112
FT NON_TER 112
FT SEQUENCE 112 AA; 13326 MW; 11086DF106524493 CRC64;
SQ

```

```

Query Match 5.1%; Score 7; DB 12; Length 112;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 6 QALQLLL 12  
| | | | |  
Db 101 QALQLLL 107

## RESULT 30

009682 PRELIMINARY; PRT; 112 AA.  
ID 009682  
AC 009682;  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Nonstructural protein NS2 (Fragment).  
GN NS.  
OS Influenzavirus A.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza A viruses.  
OX NCBI\_TaxID=197911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/Swine/Iowa/15/30;  
RX MEDLINE=97048087; PubMed=8892928;  
RA Guan Y., Shortridge K.F., Krauss S., Li P.H., Kawaoaka Y.,  
RA Webster R.G.;  
RT "Emergence of avian H1N1 influenza viruses in pigs in China."  
RL J. Virol. 70:8041-8046(1996).  
DR EMBL; U49484; AAB5096.1; -.  
DR InterPro: IPR000968; P1u NS2.  
DR Pfam: PF00601; P1u NS2; 1.  
FT NON\_TER 1 1  
FT NON\_TER 112 112  
SQ SEQUENCE 112 AA; 13297 MW; 7C6C5E8776C0016 CRC64;

Query Match 5.1%; Score 7; DB 12; Length 112;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLLL 12  
| | | | |  
Db 98 QALQLLL 104

## RESULT 31

009692 PRELIMINARY; PRT; 112 AA.  
ID 009692  
AC 009692;  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Nonstructural protein NS2 (Fragment).  
GN NS.  
OS Influenzavirus A.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza A viruses.  
OX NCBI\_TaxID=197911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/Equine/Prague/1/56;  
RX MEDLINE=97048087; PubMed=8892928;  
RA Guan Y., Shortridge K.F., Krauss S., Li P.H., Kawaoaka Y.,  
RA Webster R.G.;  
RT "Emergence of avian H1N1 influenza viruses in pigs in China."  
RL J. Virol. 70:8041-8046(1996).  
DR EMBL; U49483; AAB51006.1; -.  
DR InterPro: IPR000968; P1u NS2.  
DR Pfam: PF00601; P1u NS2; 1.  
FT NON\_TER 1 1  
FT NON\_TER 112 112  
SQ SEQUENCE 112 AA; 13480 MW; 45E79AF808E0B650 CRC64;

Query Match 5.1%; Score 7; DB 12; Length 112;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLLL 12  
| | | | |  
Db 98 QALQLLL 104

## RESULT 32

009688 PRELIMINARY; PRT; 112 AA.  
ID 009688  
AC 009688;  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Nonstructural protein NS2 (Fragment).  
GN NS.  
OS Influenzavirus A.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza A viruses.  
OX NCBI\_TaxID=197911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/Equine/Tennessee/5/86;  
RX MEDLINE=97048087; PubMed=8892928;  
RA Guan Y., Shortridge K.F., Krauss S., Li P.H., Kawaoaka Y.,  
RA Webster R.G.;  
RT "Emergence of avian H1N1 influenza viruses in pigs in China."  
RL J. Virol. 70:8041-8046(1996).  
DR EMBL; U49487; AAB51002.1; -.  
DR InterPro: IPR000968; P1u NS2.  
DR Pfam: PF00601; P1u NS2; 1.  
FT NON\_TER 1 1  
FT NON\_TER 112 112  
SQ SEQUENCE 112 AA; 13348 MW; 7400FF25359DB357 CRC64;

Query Match 5.1%; Score 7; DB 12; Length 112;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLLL 12  
| | | | |  
Db 98 QALQLLL 104

## RESULT 33

009686 PRELIMINARY; PRT; 112 AA.  
ID 009686  
AC 009686;  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Nonstructural protein NS2 (Fragment).  
GN NS.  
OS Influenzavirus A.  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza A viruses.  
OX NCBI\_TaxID=197911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/Equine/London/1416/73;  
RX MEDLINE=97048087; PubMed=8892928;  
RA Guan Y., Shortridge K.F., Krauss S., Li P.H., Kawaoaka Y.,  
RA Webster R.G.;  
RT "Emergence of avian H1N1 influenza viruses in pigs in China."  
RL J. Virol. 70:8041-8046(1996).  
DR EMBL; U49486; AAB51000.1; -.  
DR InterPro: IPR000968; P1u NS2.  
DR Pfam: PF00601; P1u NS2; 1.  
FT NON\_TER 1 1  
FT NON\_TER 112 112  
SQ SEQUENCE 112 AA; 13308 MW; 590285174DD7812D CRC64;

Query Match 5.1%; Score 7; DB 12; Length 112;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 QALQALL 12  
 DB 98 QALQALL 104

## RESULT 34

OBQLR3 PRELIMINARY; PRT; 113 AA.  
 AC OBQLR3;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Non structural protein 2 (Fragment).  
 GN NS2.  
 OS Influenza A virus (A/swine/Cotes d'Armor/1121/00(H1N1)).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 CC Influenza A viruses; Influenzavirus A.  
 CX NCBI\_TaxID=161497;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A/swine/Cotes d'Armor/1121/00(H1N1);  
 RX MEDLINE=21904432; PubMed=11907321;  
 RA Marozin S., Gregory V., Cameron K., Bennett M., Valette M., Aymard M.,  
 RA Font B., Barigazzi G., Lin Y., Hay A.;  
 RT "Antigenic and genetic diversity among swine influenza A H1N1 and H1N2  
 viruses in Europe.";  
 RL J. Gen. Virol. 83:735-745(2002).  
 DR EMBL; AJ344041; CAC86660.1; -.  
 DR InterPro; IPR000968; Flu\_NS2.  
 DR Pfam; PF00601; Flu\_NS2; I.  
 FT NON\_TER 113  
 SQ SEQUENCE 113 AA; 13371 MW; 01E1086180DD3FF2 CRC64;

Query Match 5.1%; Score 7; DB 12; Length 113;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 QALQALL 12  
 DB 101 QALQALL 107

## RESULT 35

OBQLR3 PRELIMINARY; PRT; 114 AA.  
 AC OBQLR3;  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Nonstructural protein NS2 (Fragment).  
 GN NS.  
 OS Influenzavirus A.  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 CC Influenza A viruses.  
 CX NCBI\_TaxID=197911;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A/Duck/Hong Kong/717/79;  
 RX MEDLINE=97048087; PubMed=8892928;  
 RA Guan Y., Shortridge K.F., Kraus S., Li P.H., Kawoka Y.,  
 RA Webster R.G.;  
 RT "Emergence of avian H1N1 influenza viruses in pigs in China.";  
 RL J. Virol. 70:8041-8046(1996).  
 DR EMBL; U49492; AAB51012.1; -.  
 DR InterPro; IPR000968; Flu\_NS2.  
 DR Pfam; PF00601; Flu\_NS2; I.  
 FT NON\_TER 114  
 SQ SEQUENCE 114 AA; 13440 MW; F49CDA3C07E2D005 CRC64;  
 Query Match 5.1%; Score 7; DB 12; Length 114;

Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 QALQALL 12  
 DB 100 QALQALL 106

## RESULT 36

OBQLZ1 PRELIMINARY; PRT; 115 AA.  
 AC OBQLZ1;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Non structural protein 2 (Fragment).  
 GN NS2.  
 OS Influenza A virus (A/swine/Cotes d'Armor/1122/00(H1N2)).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 CC Influenza A viruses; Influenzavirus A.  
 CX NCBI\_TaxID=169172;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A/swine/Cotes d'Armor/1122/00(H1N2);  
 RX MEDLINE=21904432; PubMed=11907321;  
 RA Marozin S., Gregory V., Cameron K., Bennett M., Valette M., Aymard M.,  
 RA Font B., Barigazzi G., Lin Y., Hay A.;  
 RT "Antigenic and genetic diversity among swine influenza A H1N1 and H1N2  
 viruses in Europe.";  
 RL J. Gen. Virol. 83:735-745(2002).  
 DR EMBL; AJ316046; CAC87385.1; -.  
 DR InterPro; IPR000968; Flu\_NS2.  
 DR Pfam; PF00601; Flu\_NS2; I.  
 FT NON\_TER 115  
 SQ SEQUENCE 115 AA; 13423 MW; 9A3299F97BD39A85 CRC64;

Query Match 5.1%; Score 7; DB 12; Length 115;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 QALQALL 12  
 DB 101 QALQALL 107

## RESULT 37

OBQLR3 PRELIMINARY; PRT; 115 AA.  
 AC OBQLR3;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Nonstructural protein 2 (Fragment).  
 GN NS2.  
 OS Influenza A virus (A/Chicken/Hong Kong/G9/97(H9N2)).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 CC Influenza A viruses; Influenzavirus A.  
 CX NCBI\_TaxID=97347;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A/Chicken/Hong Kong/G9/97;  
 RX MEDLINE=99362763; PubMed=10430948;  
 RA Guan Y., Shortridge K.F., Kraus S., Webster R.G.;  
 RT "Molecular characterization of H9N2 influenza viruses: were they the  
 RT doxor of the 'internal' genes of H5N1 viruses in Hong Kong?";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:9363-9367(1999).  
 DR EMBL; AF156472; AAD52934.2; -.  
 DR InterPro; IPR000968; Flu\_NS2.  
 DR Pfam; PF00601; Flu\_NS2; I.  
 FT NON\_TER 115  
 SQ SEQUENCE 115 AA; 13648 MW; CD12CB7123A30584 CRC64;

Query Match 5.1%; Score 7; DB 12; Length 115;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 QALQLL 12  
 |||||  
 Db 96 QALQLL 102

## RESULT 38

08QLT8 PRELIMINARY; PRT; 116 AA.

AC 08QLT8; 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Non structural protein 2 (Fragment).  
 GN NS2.  
 OS Influenza A virus (A/Swine/Italy/1523/98 (H3N2)).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza A viruses; Influenzavirus A.  
 OX NCBI\_TaxID=157799;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A/Swine/Italy/1523/98 (H3N2);  
 RX MEDLINE=21904432; PubMed=11907321;  
 RA Marozin S., Gregory V., Cameron K., Bennett M., Vajette M., Ayward M.,  
 RA Pont E., Barigazzi G., Lin Y., Hay A.;  
 RT "Antigenic and genetic diversity among swine influenza A H1N1 and H1N2  
 viruses in Europe.";  
 RL J. Gen. Virol. 83:735-745 (2002).  
 DR EMBL; AJ344024; CAC86661.1; -  
 DR InterPro; IPR000256; Flu\_NS1.  
 DR Pfam; PF00600; Flu\_NS1.1;  
 DR Non\_TER 116 116  
 FT NON\_TER 116 116  
 SQ SEQUENCE 116 AA; 13672 MW; BA186202C4DC64CC CRC64;  
 Query Match 5.1%; Score 7; DB 12; Length 116;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 QALQLL 12  
 |||||  
 Db 101 QALQLL 107

## RESULT 39

009703 PRELIMINARY; PRT; 116 AA.

AC 009703; 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Nonstructural protein NS2 (Fragment).  
 GN NS.  
 OS Influenzavirus A.  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza A viruses.  
 OX NCBI\_TaxID=197911;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A/Goose/Hong Kong/8/76;  
 RX MEDLINE=97048087; PubMed=8892928;  
 RA Guan Y., Shortridge K.F., Krauss S., Li P.H., Kawakita Y.,  
 RA Webster R.G.;  
 RT "Emergence of avian H1N1 influenza viruses in pigs in China.";  
 RL J. Virol. 70:8041-8046 (1996).  
 DR EMBL; U49495; AB51018.1; -  
 DR InterPro; IPR000256; Flu\_NS1.  
 DR InterPro; IPR000686; Flu\_NS2.

DR Pfam; PF00600; Flu\_NS1.1.  
 DR Pfam; PF00601; Flu\_NS2.1.  
 FT NON\_TER 116 116  
 SQ SEQUENCE 116 AA; 13711 MW; A7233200D99D5ABD CRC64;

Qy 6 QALQLL 12  
 |||||  
 Db 101 QALQLL 107

## RESULT 40

09Q0G6 PRELIMINARY; PRT; 117 AA.

AC 09Q0G6; 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Nonstructural protein 2 (Fragment).  
 GN NS2.  
 OS Influenza A virus (A/Chicken/Hong Kong/G23/97 (H9N2)).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza A viruses; Influenzavirus A.  
 OX NCBI\_TaxID=97348;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A/Chicken/Hong Kong/G23/97;  
 RX MEDLINE=99362763; PubMed=10430948;  
 RA Guan Y., Shortridge K.F., Krauss S., Webster R.G.;  
 RT "Molecular characterization of H9N2 influenza viruses: were they the  
 RT donors of the 'internal' genes of H5N1 viruses in Hong Kong?";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:9363-9367 (1999).  
 DR EMBL; AF156473; AAD52936.2; -  
 DR InterPro; IPR000968; Flu\_NS2.  
 DR Pfam; PF00601; Flu\_NS2.1;  
 DR Non\_TER 117 117  
 FT NON\_TER 117 117  
 SQ SEQUENCE 117 AA; 13863 MW; 6D46854FBC0EC832 CRC64;  
 Query Match 5.1%; Score 7; DB 12; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 QALQLL 12  
 |||||  
 Db 98 QALQLL 104

## RESULT 41

09Q0E4 PRELIMINARY; PRT; 118 AA.

AC 09Q0E4; 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Nonstructural protein 2 (Fragment).  
 GN NS2.  
 OS Influenza A virus (A/Quail/Arkansas/29209-1/93 (H9N2)).  
 OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
 OC Influenza A viruses; Influenzavirus A.  
 OX NCBI\_TaxID=97395;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A/Quail/Arkansas/29209-1/93;  
 RX MEDLINE=99362763; PubMed=10430948;  
 RA Guan Y., Shortridge K.F., Krauss S., Webster R.G.;  
 RT "Molecular characterization of H9N2 influenza viruses: were they the  
 RT donors of the 'internal' genes of H5N1 viruses in Hong Kong?";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:9363-9367 (1999).  
 DR EMBL; AF156484; AAD52958.2; -

```

DR InterPro; IPR000968; Flu_NS2.
DR Pfam; PF00601; Flu_NS2; 1.
FT NON_TER 1
FT NON_TER 118
SQ SEQUENCE 118 AA; 1399 MW; FP93BED84C12CC22 CRC64;

Query March
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 QALQALL 12
Db 99 QALQALL 105

RESULT 42
ID 009701 PRELIMINARY; PRT; 118 AA.
AC 009701;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Nonstructural protein NS2 (Fragment).
GN NS.
OS Influenzavirus A.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
CX NCBI_TaxID=197911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Duck/Hong Kong/193/77;
RX MEDLINE=97048087; PubMed=8892928;
RA Guan Y., Shortridge K.F., Kraus S., Li P.H., Kawoka Y.,
  Webster R.G.;
RT "Emergence of avian H1N1 influenza viruses in pigs in China.";
RL J. Virol. 70:8041-8046(1996).
DR EMBL; U49494; AAB51016.1; -.
DR InterPro; IPR000256; Flu_NS1.
DR InterPro; IPR000968; Flu_NS2.
DR Pfam; PF00600; Flu_NS1; 1.
DR Pfam; PF00601; Flu_NS2; 1.
FT NON_TER 118
FT NON_TER 118
SQ SEQUENCE 118 AA; 13932 MW; 10CA97F95200D9D CRC64;

Query March
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 QALQALL 12
Db 101 QALQALL 107

RESULT 43
ID 009694 PRELIMINARY; PRT; 118 AA.
AC 009694;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Nonstructural protein NS2 (Fragment).
GN NS.
OS Influenzavirus A.
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses.
CX NCBI_TaxID=197911;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Swine/Hong Kong/273/94;
RX MEDLINE=97048087; PubMed=8892928;
RA Guan Y., Shortridge K.F., Kraus S., Li P.H., Kawoka Y.,
  Webster R.G.;
RT "Emergence of avian H1N1 influenza viruses in pigs in China.";

```

```

RL J. Virol. 70:8041-8046(1996).
DR EMBL; U49490; AAB51008.1; -.
DR InterPro; IPR000968; Flu_NS2.
DR Pfam; PF00601; Flu_NS2; 1.
FT NON_TER 1
FT NON_TER 118
SQ SEQUENCE 118 AA; 14045 MW; B5A284926501A684 CRC64;

Query March
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 QALQALL 12
Db 100 QALQALL 106

RESULT 44
ID 090064 PRELIMINARY; PRT; 118 AA.
AC 090064;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NS2 (Fragment).
GN NS2.
OS Influenza A virus (A/Pigeon/Hong Kong/Y233/97(H9N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
CX NCBI_TaxID=97349;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Pigeon/Hong Kong/Y233/97;
RX MEDLINE=99362763; PubMed=10430948;
RA Guan Y., Shortridge K.F., Kraus S., Webster R.G.;
RT "Molecular characterization of H9N2 influenza viruses: were they the
  donors of the 'internal' genes of H5N1 viruses in Hong Kong?";
RL Proc. Natl. Acad. Sci. U.S.A. 96:9363-9367(1999).
DR EMBL; AF156474; AAD52938.2; -.
DR InterPro; IPR000968; Flu_NS2.
DR Pfam; PF00601; Flu_NS2; 1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 118 AA; 13950 MW; 6D46851C8C3237A2 CRC64;

Query March
Best Local Similarity 100.0%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 QALQALL 12
Db 99 QALQALL 105

RESULT 45
ID 090062 PRELIMINARY; PRT; 119 AA.
AC 090062;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NS2 (Fragment).
GN NS2.
OS Influenza A virus (A/Chicken/Beijing/1/94(H9N2)).
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
CX NCBI_TaxID=97354;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A/Chicken/Beijing/1/94;
RX MEDLINE=99362763; PubMed=10430948;
RA Guan Y., Shortridge K.F., Kraus S., Webster R.G.;
RT "Molecular characterization of H9N2 influenza viruses: were they the

```

RT donors of the 'internal' genes of HSN1 viruses in Hong Kong?";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:9363-9367(1999).  
DR EMBL; AF156480; AAD52950.2; -.  
DR Interpro; IPR000968; Flu\_NS2.  
DR Pfam; PF00601; Flu\_NS2; I.  
FT NON\_TER 1 1  
FT NON\_TER 119 119  
SQ SEQUENCE 119 AA; 14066 MW; 7426887CF83AAPDB CRC64;

Query Match 5.1%; Score 7; DB 12; Length 119;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLLL 12  
Db 100 QALQLLL 106

RESULT 46  
Q9INP2 PRELIMINARY; PRT; 119 AA.  
AC Q9INP2;  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE Nonstructural protein 2.  
GN NS2.  
OS Influenza A virus (A/Hong Kong/481/97(HSN1)).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza A viruses; Influenzavirus A.  
OX NCBI\_TaxID=107423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/Hong Kong/481/97;  
RX MEDLINE=20231940; PubMed=10769072;  
RA Hironaka Y., Yamazaki Y., Fukushima T., Saito T., Lindstrom S.E.,  
RA Onoe K., Nerome R., Lim W., Sugita S., Nerome K.;  
RT "Evolutionary Characterization of the six internal genes of HSN1 human  
RT Influenza A virus";  
RL J. Gen. Virol. 81:1293-1303(2000).  
DR EMBL; AF115288; AAF75116.1; -.  
DR Interpro; IPR000968; Flu\_NS2.  
DR Pfam; PF00601; Flu\_NS2; I.  
SQ SEQUENCE 119 AA; 14088 MW; B0CA2498B76524E2 CRC64;  
FT NON\_TER 1 1  
FT NON\_TER 119 119  
SQ SEQUENCE 119 AA; 14088 MW; B0CA2498B76524E2 CRC64;  
Query Match 5.1%; Score 7; DB 12; Length 119;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLLL 12  
Db 99 QALQLLL 105

RESULT 47  
Q9QOF8 PRELIMINARY; PRT; 119 AA.  
AC Q9QOF8;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Nonstructural protein 2 (Fragment).  
GN NS2.  
OS Influenza A virus (A/Quail/Hong Kong/G1/97 (H9N2)).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza A viruses; Influenzavirus A.  
OX NCBI\_TaxID=97352;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/Quail/Hong Kong/G1/97;  
RX MEDLINE=99362763; PubMed=10430948;  
RA Guan Y., Shortridge K.F., Krauss S., Webster R.G.;  
RT "Molecular characterization of H9N2 influenza viruses: were they the

RT donors of the 'internal' genes of HSN1 viruses in Hong Kong?";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:9363-9367(1999).  
DR EMBL; AF156477; AAD52944.2; -.  
DR Interpro; IPR000968; Flu\_NS2.  
DR Pfam; PF00601; Flu\_NS2; I.  
FT NON\_TER 1 1  
FT NON\_TER 119 119  
SQ SEQUENCE 119 AA; 14119 MW; D42710049DE8932A CRC64;

Query Match 5.1%; Score 7; DB 12; Length 119;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLLL 12  
Db 100 QALQLLL 106

RESULT 48  
Q9QOF4 PRELIMINARY; PRT; 119 AA.  
AC Q9QOF4;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Nonstructural protein 2 (Fragment).  
GN NS2.  
OS Influenza A virus (A/Quail/Hong Kong/AF157/93(H9N2)).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza A viruses; Influenzavirus A.  
OX NCBI\_TaxID=106422;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/Quail/Hong Kong/AF157/92;  
RX MEDLINE=99362763; PubMed=10430948;  
RA Guan Y., Shortridge K.F., Krauss S., Webster R.G.;  
RT "Molecular characterization of H9N2 influenza viruses: were they the  
RT donors of the 'internal' genes of HSN1 viruses in Hong Kong?";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:9363-9367(1999).  
DR EMBL; AF156479; AAD52948.2; -.  
DR Interpro; IPR000968; Flu\_NS2.  
DR Pfam; PF00601; Flu\_NS2; I.  
FT NON\_TER 1 1  
FT NON\_TER 119 119  
SQ SEQUENCE 119 AA; 14052 MW; EDB6ED72FC300F7E CRC64;  
FT NON\_TER 1 1  
FT NON\_TER 119 119  
SQ SEQUENCE 119 AA; 14052 MW; EDB6ED72FC300F7E CRC64;  
Query Match 5.1%; Score 7; DB 12; Length 119;  
Best Local Similarity 100.0%; Pred. No. 20;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLLL 12  
Db 100 QALQLLL 106

RESULT 49  
Q9QOG2 PRELIMINARY; PRT; 119 AA.  
AC Q9QOG2;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Nonstructural protein 2 (Fragment).  
GN NS2.  
OS Influenza A virus (A/Duck/Hong Kong/Y280/97(H9N2)).  
OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;  
OC Influenza A viruses; Influenzavirus A.  
OX NCBI\_TaxID=97350;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A/Duck/Hong Kong/Y280/97;  
RX MEDLINE=99362763; PubMed=10430948;  
RA Guan Y., Shortridge K.F., Krauss S., Webster R.G.;

RT "Molecular characterization of H9N2 influenza viruses: were they the  
 RT donors of the 'internal' genes of H5N1 viruses in Hong Kong?";

RU Proc. Natl. Acad. Sci. U.S.A. 96:9363-9367(1999).

DR EMBL; AF156475; AAD52940.2; -.

DR InterPro; IPR000968; Flu\_NS2.

DR Pfam; PF00601; Flu\_NS2; 1.

FT NON\_TER 1 1

FT NON\_TER 119 119

SEQUENCE 119 AA; 14135 MW; 7596887CF93BAFDA CRC64;

Query Match

Best Local Similarity 5.1%; Score 7; DB 12; Length 119;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLL 12

Db 100 QALQLL 106

RESULT 50

Q9Q0F0 PRELIMINARY; PRT; 119 AA.

AC Q9Q0F0;

DT 01-MAY-2000 (TRENBLREL. 13, Created)

DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)

DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)

DE Nonstructural protein 2 (Fragment).

GN NS2.

OS Influenza A virus (A/Chicken/Korea/38349-p96323/96 (H9N2)).

OC Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;

OC Influenza A viruses; Influenzavirus A.

OX NCBI\_Taxid=97358;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A/Chicken/Korea/38349-P96323/96;

RA MEDLINE=9362763; PubMed=10430948;

RA Guan Y., Shortridge K.F., Kraus S., Webster R.G.;

RT "Molecular characterization of H9N2 influenza viruses: were they the

RT donors of the 'internal' genes of H5N1 viruses in Hong Kong?";

RU Proc. Natl. Acad. Sci. U.S.A. 96:9363-9367(1999).

DR EMBL; AF156481; AAD52952.2; -.

DR InterPro; IPR000968; Flu\_NS2.

DR Pfam; PF00601; Flu\_NS2; 1.

FT NON\_TER 1 1

FT NON\_TER 119 119

SEQUENCE 119 AA; 14064 MW; 2CAC829D1323117F CRC64;

Query Match

Best Local Similarity 5.1%; Score 7; DB 12; Length 119;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 QALQLL 12

Db 100 QALQLL 106

Search completed: January 21, 2004, 12:27:23

Job time : 38 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 12:20:56 ; Search time 41 Seconds  
(without alignments)  
530.379 Million cell updates/sec

Title: US-09-941-314-2

Perfect score: 137  
Sequence: 1 MABPMQALDQLLALTLTMA.....VFAVPWFQYKIKKSCSSD 137

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Database :

A\_Geneseq\_19Jun03:\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	137	100.0	137	23	AAU79852 Human cystatin-8
2	117	85.4	117	23	AAU79854 Human cystatin-8
3	115	83.9	115	23	AAU79853 Human cystatin-8
4	80	58.4	80	23	AAU79865 Human cystatin-8
5	59	43.1	59	23	AAU79866 Human cystatin-8
6	52	38.0	52	23	AAU79864 Human cystatin-8
7	50	36.5	50	22	ABG48915 Human liver peptid
8	50	36.5	50	22	ABR28913 Peptide #1564 enco
9	50	36.5	50	22	ABR32369 Peptide #5040 enco

10	50	36.5	50	22	ABR34086 Peptide #1592 enco
11	50	36.5	50	22	ABR19524 Protein #1523 enco
12	50	36.5	50	22	AAW54871 Human brain expres
13	50	36.5	50	22	AAW67252 Human bone marrow
14	50	36.5	50	22	AAW15096 Peptide #1530 enco
15	50	36.5	50	22	AAW27545 Peptide #1582 enco
16	50	36.5	50	22	AAW02833 Peptide #1515 enco
17	50	36.5	50	22	ABG36903 Human peptid enco
18	49	35.8	49	23	AAU79863 Human cystatin-8
19	48	35.0	48	23	AAU79867 Human cystatin-8
20	46	33.6	46	23	AAU79860 Human cystatin-8
21	36	26.3	36	23	AAU79856 Human cystatin-8
22	35	25.5	35	23	AAU79858 Human cystatin-8
23	33	24.1	33	23	AAU79862 Human cystatin-8
24	27	19.7	27	23	AAU79859 Human cystatin-8
25	24	17.5	24	23	AAU79861 Human cystatin-8
26	22	16.1	22	23	AAU79855 Human cystatin-8
27	22	13.1	18	23	AAU79857 Human cystatin-8
28	8	5.8	1252	22	ABR59363 Drosophila melanog
29	7	5.1	60	22	AAU68887 Propionibacterium
30	7	5.1	97	22	ABG03232 Novel human diagno
31	7	5.1	98	21	AAW21482 Arabidopsis thalia
32	7	5.1	98	21	AAW48767 Arabidopsis thalia
33	7	5.1	100	21	AAW21481 Arabidopsis thalia
34	7	5.1	100	21	AAW48766 Arabidopsis thalia
35	7	5.1	102	21	AAW42196 Arabidopsis thalia
36	7	5.1	112	21	AAW83235 Novel secreted pro
37	7	5.1	112	21	AAW42195 Arabidopsis thalia
38	7	5.1	182	23	ABR89858 Human polyepitide
39	7	5.1	184	23	ABR90713 Chlamydia pneumoni
40	7	5.1	228	22	ABR68742 Drosophila melanog
41	7	5.1	267	22	ABR70167 Drosophila melanog
42	7	5.1	288	23	ABR53808 Lactococcus lactis
43	7	5.1	330	21	AAW15904 E. coli proliferat
44	7	5.1	333	21	AAW22146 Ramoplanin biosynt
45	7	5.1	361	22	ABR69595 Drosophila melanog
46	7	5.1	381	22	AAU08679 Human PCTP3a poly
47	7	5.1	381	22	AAW61141 Human NOV11 protei
48	7	5.1	394	20	AAW37807 Amino acid sequenc
49	7	5.1	415	22	ABR36700 Human tumor necro
50	7	5.1	415	24	ABR96138 Mouse lymphotoxin-
51	7	5.1	429	22	AAU00392 Human secreted pro
52	7	5.1	429	23	ABR49892 Listeria monocytog
53	7	5.1	502	23	ABR30004 Streptococcus poly
54	7	5.1	506	24	ABR02352 S. pneumoniae type
55	7	5.1	507	21	AAW81580 Streptococcus pneu
56	7	5.1	519	22	AAW80928 Ketocacyl ACP synth
57	7	5.1	650	21	AAW82527 Human RPI05 protei
58	7	5.1	651	19	AAW28510 Product of clone J
59	7	5.1	661	19	AAW47274 Human B-cell activ
60	7	5.1	661	20	AAW87556 B cell surface pro
61	7	5.1	741	22	ABG25060 Novel human diagno
62	7	5.1	741	22	ABG25060 Drosophila melanog
63	7	5.1	848	22	ABG92087 Drosophila melanog
64	7	5.1	865	22	ABG03234 Phototransduc lumini
65	7	5.1	878	23	ABR93298 Human hexokinase 5
66	7	5.1	917	23	ABR04582 Drosophila melanog
67	7	5.1	995	22	ABR58605 Drosophila melanog
68	7	5.1	1815	22	ABR68268 Phototransduc lumini
69	7	5.1	1844	18	AAW18302 Toxin TcdA11, enco
70	7	5.1	1844	19	AAW56558 Drosophila melanog
71	7	5.1	2404	22	ABR69209 Phototransduc lumini
72	7	5.1	2504	18	AAW17871 Toxin TcdA, encode
73	7	5.1	2504	19	AAW56557 Phototransduc tcdA
74	7	5.1	2504	22	AAW72610 P. luminescens (N-
75	7	5.1	2504	23	ABG32654

## ALIGNMENTS

RESULT 1  
AAU79852

ID AAV79852 standard; Protein; 137 AA.  
 XX AAV79852;  
 AC  
 XX 15-JUL-2002 (first entry)  
 DT  
 XX Human cystatin-8 (Zcys8).  
 DE  
 XX Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;  
 KM spermatogenesis; seminal fluid viscosity; cryopreserved sperm;  
 KM sperm motility; fertilisation.  
 XX Homo sapiens.  
 OS  
 XX MO200220567-A2.  
 PN  
 XX 14-MAR-2002.  
 PD  
 XX 29-AUG-2001; 2001WO-US26868.  
 PF  
 XX 01-SEP-2000; 2000US-230230P.  
 PR  
 XX (ZYMO) ZYMOGENETICS INC.  
 PA  
 PI Holloway JL, Gao Z, Bishop PD;  
 XX  
 XX WPI; 2002-383044/41.  
 DR N-PSDB; ABK49522.  
 DR  
 PT Novel isolated mammalian cystatin-8 polypeptide useful for promoting  
 PT spermatogenesis, and inhibiting cancer procoagulant protein which leads  
 PT to inhibition of thrombotic events associated with cancer -  
 XX  
 PS Claim 2; Page 93-94; 100pp; English.  
 XX  
 XX The invention describes an isolated mammalian cystatin-8 (Zcys8)  
 CC polypeptide (I). (I) is useful for: inhibiting cancer procoagulant  
 CC protein in an individual and thus inhibiting the thrombotic events  
 CC associated with cancer; promoting spermatogenesis; modulating seminal  
 CC fluid viscosity; enhancing viability of cryopreserved sperm; sperm  
 CC motility and fertilisation; and as antigenic peptides to generate  
 CC antibodies. Zcys8 is useful as research reagent for characterising sites  
 CC of interaction between Zcys8 and its receptor. Zcys8 is useful in  
 CC enhancing fertilisation during assisted reproduction in humans and in  
 CC animals. Anti-(I) antibodies are useful to screen biological samples like  
 CC blood, urine, saliva, tissue biopsy and autopsy material in vitro for the  
 CC presence of Zcys8. The antibodies are also useful to isolate large  
 CC quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.  
 CC The polynucleotide encoding (I) is useful to detect and to localise the  
 CC expression of a Zcys8 gene in a biological sample and Zcys8  
 CC oligonucleotide probes are useful for in vivo diagnosis. The  
 CC polynucleotide encoding (I) is useful in determining whether a subject's  
 CC chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene  
 CC copy number changes, insertions, deletions, restriction site changes and  
 CC rearrangements and genetic alterations that inactivate the Zcys8 gene.  
 CC This is the amino acid sequence of human cystatin-8 (Zcys8).  
 XX  
 XX Sequence 137 AA;  
 SQ  
 Query Match 100.0%; Score 137; DB 23; Length 137;  
 Best Local Similarity 100.0%; Pred. No. 8,6e-132;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 121 VPMFEQYKINKSCSSD 137  
 RESULT 2  
 ID AAV79854 standard; Protein; 117 AA.  
 XX AAV79854;  
 AC  
 XX 15-JUL-2002 (first entry)  
 DT  
 XX Human cystatin-8 (Zcys8) antigenic fragment #2.  
 DE  
 XX Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;  
 KM spermatogenesis; seminal fluid viscosity; cryopreserved sperm;  
 KM sperm motility; fertilisation; antigenic fragment.  
 XX Homo sapiens.  
 OS  
 XX MO200220567-A2.  
 PN  
 XX 14-MAR-2002.  
 PD  
 XX 29-AUG-2001; 2001WO-US26868.  
 PF  
 XX 01-SEP-2000; 2000US-230230P.  
 PR  
 XX (ZYMO) ZYMOGENETICS INC.  
 PA  
 PI Holloway JL, Gao Z, Bishop PD;  
 XX  
 XX WPI; 2002-383044/41.  
 DR  
 PT Novel isolated mammalian cystatin-8 polypeptide useful for promoting  
 PT spermatogenesis, and inhibiting cancer procoagulant protein which leads  
 PT to inhibition of thrombotic events associated with cancer -  
 XX  
 PS Claim 2; Page 94-95; 100pp; English.  
 XX  
 XX The invention describes an isolated mammalian cystatin-8 (Zcys8)  
 CC polypeptide (I). (I) is useful for: inhibiting cancer procoagulant  
 CC protein in an individual and thus inhibiting the thrombotic events  
 CC associated with cancer; promoting spermatogenesis; modulating seminal  
 CC fluid viscosity; enhancing viability of cryopreserved sperm; sperm  
 CC motility and fertilisation; and as antigenic peptides to generate  
 CC antibodies. Zcys8 is useful as research reagent for characterising sites  
 CC of interaction between Zcys8 and its receptor. Zcys8 is useful in  
 CC enhancing fertilisation during assisted reproduction in humans and in  
 CC animals. Anti-(I) antibodies are useful to screen biological samples like  
 CC blood, urine, saliva, tissue biopsy and autopsy material in vitro for the  
 CC presence of Zcys8. The antibodies are also useful to isolate large  
 CC quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.  
 CC The polynucleotide encoding (I) is useful to detect and to localise the  
 CC expression of a Zcys8 gene in a biological sample and Zcys8  
 CC oligonucleotide probes are useful for in vivo diagnosis. The  
 CC polynucleotide encoding (I) is useful in determining whether a subject's  
 CC chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene  
 CC copy number changes, insertions, deletions, restriction site changes and  
 CC rearrangements and genetic alterations that inactivate the Zcys8 gene.  
 CC This sequence represents an antigenic fragment of human cystatin-8  
 CC (Zcys8).  
 XX  
 XX Sequence 117 AA;  
 SQ  
 Query Match 85.4%; Score 117; DB 23; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 2e-111;  
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 61 EYHNVEMQWTTCCQKPEITNCVPOERBLHKOVNCFPSVFAVPMFBOYKILINKSCSSD 117

## RESULT 3

ID AAU79853 standard; Protein; 115 AA.

AC AAU79853;

DT 15-JUL-2002 (first entry)

DE Human cystatin-8 (Zcys8) antigenic fragment #1.

XX Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;  
 XX spermatogenesis; seminal fluid viscosity; cryopreserved sperm;  
 XX sperm motility; fertilisation; antigenic fragment.

OS Homo sapiens.

PN WO200220567-A2.

PD 14-MAR-2002.

PF 29-AUG-2001; 2001MO-US26868.

PR 01-SEP-2000; 2000US-230230P.

PR (ZYMO) ZYMOGENETICS INC.

PI Holloway JL, Gao Z, Bishop PD;

PI WPI; 2002-383044/41.

PT Novel isolated mammalian cystatin-8 polypeptide useful for promoting  
 PT spermatogenesis, and inhibiting cancer procoagulant protein which leads  
 PT to inhibition of thrombotic events associated with cancer

PS Claim 2; Page 94; 100pp; English.

XX The invention describes an isolated mammalian cystatin-8 (Zcys8)  
 CC polypeptide (I). (I) is useful for: inhibiting cancer procoagulant  
 CC protein in an individual and thus inhibiting the thrombotic events  
 CC associated with cancer; promoting spermatogenesis, modulating seminal  
 CC fluid viscosity, enhancing viability of cryopreserved sperm, sperm  
 CC motility and fertilisation; and as antigenic peptides to generate  
 CC antibodies. Zcys8 is useful as research reagent for characterising sites  
 CC of interaction between Zcys8 and its receptor. Zcys8 is useful in  
 CC enhancing fertilisation during assisted reproduction in humans and in  
 CC animals. Anti-(I) antibodies are useful to screen biological samples like  
 CC blood, urine, saliva, tissue biopsy and autopsy material in vitro for the  
 CC presence of Zcys8. The antibodies are also useful to isolate large  
 CC quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.  
 CC The polynucleotide encoding (I) is useful to detect and to localise the  
 CC expression of a Zcys8 gene in a biological sample and Zcys8  
 CC oligonucleotide probes are useful for in vivo diagnosis. The  
 CC polynucleotide encoding (I) is useful in determining whether a subject's  
 CC chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene  
 CC copy number changes, insertions, deletions, restriction site changes and  
 CC rearrangements and genetic alterations that inactivate the Zcys8 gene.  
 CC This sequence represents an antigenic fragment of human cystatin-8  
 CC (Zcys8).

SO Sequence 115 AA;

Query Match 83.9%; Score 115; DB 23; Length 115;

Best Local Similarity 100.0%; Pred. No. 2.1e-109; Indels 0; Gaps 0;

Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 YQARKTFPSVHVMVAENVAKDSLQWITPOYNKESDCKHPIFRYLKQROVTDHLEY 82  
 DB 1 YQARKTFPSVHVMVAENVAKDSLQWITPOYNKESDCKHPIFRYLKQROVTDHLEY 60

QY 83 HLNVMQWTTCCQKPEITNCVPOERBLHKOVNCFPSVFAVPMFBOYKILINKSCSSD 137  
 DB 61 HLNVMQWTTCCQKPEITNCVPOERBLHKOVNCFPSVFAVPMFBOYKILINKSCSSD 115

## RESULT 4

ID AAU79865 standard; Peptide; 80 AA.

AC AAU79865;

DT 15-JUL-2002 (first entry)

DE Human cystatin-8 (Zcys8) antigenic fragment #13.

XX Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;  
 XX spermatogenesis; seminal fluid viscosity; cryopreserved sperm;  
 XX sperm motility; fertilisation; antigenic peptide.

OS Homo sapiens.

PN WO200220567-A2.

PD 14-MAR-2002.

PF 29-AUG-2001; 2001MO-US26868.

PR 01-SEP-2000; 2000US-230230P.

PR (ZYMO) ZYMOGENETICS INC.

PI Holloway JL, Gao Z, Bishop PD;

PI WPI; 2002-383044/41.

PT Novel isolated mammalian cystatin-8 polypeptide useful for promoting  
 PT spermatogenesis, and inhibiting cancer procoagulant protein which leads  
 PT to inhibition of thrombotic events associated with cancer

PS Claim 2; Page 98; 100pp; English.

XX The invention describes an isolated mammalian cystatin-8 (Zcys8)  
 CC polypeptide (I). (I) is useful for: inhibiting cancer procoagulant  
 CC protein in an individual and thus inhibiting the thrombotic events  
 CC associated with cancer; promoting spermatogenesis, modulating seminal  
 CC fluid viscosity, enhancing viability of cryopreserved sperm, sperm  
 CC motility and fertilisation; and as antigenic peptides to generate  
 CC antibodies. Zcys8 is useful as research reagent for characterising sites  
 CC of interaction between Zcys8 and its receptor. Zcys8 is useful in  
 CC enhancing fertilisation during assisted reproduction in humans and in  
 CC animals. Anti-(I) antibodies are useful to screen biological samples like  
 CC blood, urine, saliva, tissue biopsy and autopsy material in vitro for the  
 CC presence of Zcys8. The antibodies are also useful to isolate large  
 CC quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.  
 CC The polynucleotide encoding (I) is useful to detect and to localise the  
 CC expression of a Zcys8 gene in a biological sample and Zcys8  
 CC oligonucleotide probes are useful for in vivo diagnosis. The  
 CC polynucleotide encoding (I) is useful in determining whether a subject's  
 CC chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene  
 CC copy number changes, insertions, deletions, restriction site changes and  
 CC rearrangements and genetic alterations that inactivate the Zcys8 gene.  
 CC This sequence represents an antigenic peptide of human cystatin-8  
 CC (Zcys8).

SO Sequence 80 AA;

Query Match 58.4%; Score 80; DB 23; Length 80;

Best Local Similarity 100.0%; Pred. No. 8.2e-74; Indels 0; Gaps 0;

Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 53 QYNKESDDCKHPIFRYLKQROVTDHLEHNLVEMQWTTCCQKPEITNCVPOERBLHKOV 112  
 DB 1 QYNKESDDCKHPIFRYLKQROVTDHLEHNLVEMQWTTCCQKPEITNCVPOERBLHKOV 60

QY 113 NCFPSVFAVPMFEQYKIINK 132  
DB 61 NCFPSVFAVPMFEQYKIINK 80

RESULT 5  
AAU79866  
AAU79866 standard; Peptide; 59 AA.

AAU79866;

15-JUL-2002 (first entry)

Human cystatin-8 (Zcys8) antigenic fragment #14.

Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;  
spermatogenesis; seminal fluid viscosity; cryopreserved sperm;  
sperm motility; fertilisation; antigenic peptide.

Homo sapiens.

WO200220567-A2.

14-MAR-2002.

29-AUG-2001; 2001WO-US26868.

01-SEP-2000; 2000US-230230P.

(ZYMO) ZYMOGENETICS INC.

Holloway JL, Gao Z, Bishop PD;

WPI; 2002-383044/41.

Novel isolated mammalian cystatin-8 polypeptide useful for promoting  
spermatogenesis, and inhibiting cancer procoagulant protein which leads  
to inhibition of thrombotic events associated with cancer

Claim 2; Page 99; 100pp; English.

The invention describes an isolated mammalian cystatin-8 (Zcys8)  
polypeptide (I). (I) is useful for: inhibiting cancer procoagulant  
protein in an individual and thus inhibiting the thrombotic events  
associated with cancer; promoting spermatogenesis; modulating seminal  
fluid viscosity; enhancing viability of cryopreserved sperm; sperm  
motility and fertilisation; and as antigenic peptides to generate  
antibodies. Zcys8 is useful as research reagent for characterising sites  
of interaction between Zcys8 and its receptor. Zcys8 is useful in  
enhancing fertilisation during assisted reproduction in humans and in  
animals. Anti-(I) antibodies are useful to screen biological samples like  
blood, urine, saliva, tissue biopsy and autopsy material in vitro for the  
presence of Zcys8. The antibodies are also useful to isolate large  
quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.  
The polynucleotide encoding (I) is useful to detect and to localise the  
expression of a Zcys8 gene in a biological sample and Zcys8  
oligonucleotide probes are useful for in vivo diagnosis. The  
polynucleotide encoding (I) is useful in determining whether a subject's  
chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene  
copy number changes, insertions, deletions, restriction site changes and  
rearrangements and genetic alterations that inactivate the Zcys8 gene.  
This sequence represents an antigenic peptide of human cystatin-8  
(Zcys8).

Sequence 59 AA;

Query Match 43.1%; Score 59; DB 23; Length 59;  
Best Local Similarity 100.0%; Pred. No. 1.7e-52;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

74 RQVTDHLEHNLAVMOMTTCQKPEFTNCVPERELHKNQVCFPSVFAVPMFEQYKIINK 132  
|||||

DB 1 RQVTDHLEHNLAVMOMTTCQKPEFTNCVPERELHKNQVCFPSVFAVPMFEQYKIINK 59

RESULT 6  
AAU79864  
AAU79864 standard; Peptide; 52 AA.

AAU79864;

15-JUL-2002 (first entry)

Human cystatin-8 (Zcys8) antigenic fragment #12.

Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;  
spermatogenesis; seminal fluid viscosity; cryopreserved sperm;  
sperm motility; fertilisation; antigenic peptide.

Homo sapiens.

WO200220567-A2.

14-MAR-2002.

29-AUG-2001; 2001WO-US26868.

01-SEP-2000; 2000US-230230P.

(ZYMO) ZYMOGENETICS INC.

Holloway JL, Gao Z, Bishop PD;

WPI; 2002-383044/41.

Novel isolated mammalian cystatin-8 polypeptide useful for promoting  
spermatogenesis, and inhibiting cancer procoagulant protein which leads  
to inhibition of thrombotic events associated with cancer

Claim 2; Page 98; 100pp; English.

The invention describes an isolated mammalian cystatin-8 (Zcys8)  
polypeptide (I). (I) is useful for: inhibiting cancer procoagulant  
protein in an individual and thus inhibiting the thrombotic events  
associated with cancer; promoting spermatogenesis; modulating seminal  
fluid viscosity; enhancing viability of cryopreserved sperm; sperm  
motility and fertilisation; and as antigenic peptides to generate  
antibodies. Zcys8 is useful as research reagent for characterising sites  
of interaction between Zcys8 and its receptor. Zcys8 is useful in  
enhancing fertilisation during assisted reproduction in humans and in  
animals. Anti-(I) antibodies are useful to screen biological samples like  
blood, urine, saliva, tissue biopsy and autopsy material in vitro for the  
presence of Zcys8. The antibodies are also useful to isolate large  
quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.  
The polynucleotide encoding (I) is useful to detect and to localise the  
expression of a Zcys8 gene in a biological sample and Zcys8  
oligonucleotide probes are useful for in vivo diagnosis. The  
polynucleotide encoding (I) is useful in determining whether a subject's  
chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene  
copy number changes, insertions, deletions, restriction site changes and  
rearrangements and genetic alterations that inactivate the Zcys8 gene.  
This sequence represents an antigenic peptide of human cystatin-8  
(Zcys8).

Sequence 52 AA;

Query Match 38.0%; Score 52; DB 23; Length 52;  
Best Local Similarity 100.0%; Pred. No. 2.2e-45;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 KESDDKXHFRIFRVLKYQROVTDHLEHNLAVMOMTTCQKPEFTNCVPERE 107  
DB 1 KESDDKXHFRIFRVLKYQROVTDHLEHNLAVMOMTTCQKPEFTNCVPERE 52  
|||||

## RESULT 7

ABG48915

ID ABG48915 standard; Peptide; 50 AA.

AC ABG48915;

XX 25-FEB-2003 (first entry)

DE Human liver peptide, SEQ ID No 27563.

XX Human liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;

XX hypercholesterolaemia; coronary heart disease.

XX Homo sapiens.

XX WO200157273-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00664.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLB-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-468898/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analysing gene expression in human adult liver -

XX Claim 27; SEQ ID No 27563; 658bp; English.

CC The invention relates to a single exon nucleic acid probe (SENP) (I) for  
 CC measuring human gene expression in a sample derived from human adult  
 CC liver, comprising one of 13109 defined nucleotide sequences given in the  
 CC specification (or complements/ fragments). The probe hybridises at high  
 CC stringency to a nucleic acid molecule expressed in the human adult  
 CC liver. (I) may be used for predicting, measuring and displaying gene  
 CC expression in samples derived from human adult liver. The genes  
 CC identified may be involved in genetic liver diseases such as cirrhosis,  
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which  
 CC is associated with coronary heart disease. ABG47348-ABG5930 represent  
 CC human liver single exon encoded peptides of the invention.  
 CC Note: The sequence information for this patent does not appear in the  
 CC printed specification but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 50 AA;

Query Match 36.5%; Score 50; DB 22; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-43; Indels 0; Gaps 0;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 RKTFTLSVHEVMAVENTAKDSLQWITDQYNKESDCKHFRIFRVLKQVRQ 75  
 DB 1 RKTFTLSVHEVMAVENTAKDSLQWITDQYNKESDCKHFRIFRVLKQVRQ 50

## RESULT 8

ABB28913

ID ABB28913 standard; Peptide; 50 AA.

XX ABB28913;

XX 01-FEB-2002 (first entry)

XX Peptide #1564 encoded by breast cell single exon nucleic acid probe.

DE Human; microarray; single exon probe; gene expression; breast;

XX disease; cancer.

XX Homo sapiens.

XX WO200157271-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00662.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLB-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-466933/54.

XX New spatially-addressable set of single exon nucleic acid probes,

XX useful for measuring gene expression in sample derived from human

XX breast, comprises number of single exon nucleic acid probes -

XX Claim 27; SEQ ID No 11881; 327bp + sequence listing; English.

CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human breast and BT 474 cells. The method involves contacting  
 CC the probes with a collection of detectably labelled nucleic acids  
 CC derived from mRNA of human breast, and then measuring the label  
 CC bound to each probe of the microarray. The probes are useful for  
 CC verifying the expression of regions of genomic DNA predicted to  
 CC encode proteins. They are useful for gene discovery, and for  
 CC determining predisposition and/or prognosing breast disease. Gene  
 CC expression analysis is useful for assessing the toxicity of chemical  
 CC agents on cells. The microarray of this invention presents a far greater  
 CC diversity of probes for measuring gene expression, with far less bias  
 CC than expressed sequence tag microarrays. The method is suitable for  
 CC rapid production of functional information from genomic sequence. The  
 CC present sequence is a peptide encoded by a single exon nucleic acid  
 CC probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 50 AA;

Query Match 36.5%; Score 50; DB 22; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-43;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 26 RKTFTLSVHEVMAVENTAKDSLQWITDQYNKESDCKHFRIFRVLKQVRQ 75  
 DB 1 RKTFTLSVHEVMAVENTAKDSLQWITDQYNKESDCKHFRIFRVLKQVRQ 50

## RESULT 9

ABB32389

ID ABB32389 standard; Peptide; 50 AA.

XX ABB32389;

XX 01-FEB-2002 (first entry)

DE Peptide #5040 encoded by breast cell single exon nucleic acid probe.  
 XX Human; microarray; single exon probe; gene expression; breast;  
 KW disease; cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157271-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00662.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 XX 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-496933/54.  
 XX  
 PT New spatially-addressable set of single exon nucleic acid probes,  
 PT useful for measuring gene expression in sample derived from human  
 PT breast, comprises number of single exon nucleic acid probes -  
 XX  
 PS Claim 27; SEQ ID NO 15357; 327bp + sequence listing; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human breast and BT 474 cells. The method involves contacting  
 CC the probes with a collection of detectably labelled nucleic acids  
 CC derived from mRNA of human breast, and then measuring the label  
 CC bound to each probe of the microarray. The probes are useful for  
 CC verifying the expression of regions of genomic DNA predicted to  
 CC encode proteins. They are useful for gene discovery, and for  
 CC determining predisposition and/or prognosing breast disease. Gene  
 CC expression analysis is useful for assessing the toxicity of chemical  
 CC agents on cells. The microarray of this invention presents a far greater  
 CC diversity of probes for measuring gene expression, with far less bias  
 CC than expressed sequence tag microarrays. The method is suitable for  
 CC rapid production of functional information from genomic sequence. The  
 CC present sequence is a peptide encoded by a single exon nucleic acid  
 CC probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 50 AA;  
 XX  
 Query Match 36.5%; Score 50; DB 22; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-43;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 26 RKKTFLSVHEVMAVENYAKDSLQWITDQYNKESDPDKYHFRIFRYLKVQRQ 75  
 Db 1 RKKTFLSVHEVMAVENYAKDSLQWITDQYNKESDPDKYHFRIFRYLKVQRQ 50  
 XX  
 RESULT 10  
 ABB34086  
 ID ABB34086 standard; Peptide; 50 AA.  
 XX  
 AC ABB34086;  
 XX  
 DT 04-FEB-2002 (first entry)  
 XX  
 DE Peptide #1592 encoded by human foetal liver single exon probe.

XX  
 KW Human; foetal liver; gene expression; single exon nucleic acid probe.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157277-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00669.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 XX 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-483447/52.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human fetal liver -  
 XX  
 PS Claim 27; SEQ ID NO 26721; 639bp + sequence listing; English.  
 XX  
 CC The invention relates to a single exon nucleic acid probe for  
 CC measuring human gene expression in a sample derived from human foetal  
 CC liver. The single exon nucleic acid probes may be used for predicting,  
 CC measuring and displaying gene expression in samples derived from human  
 CC foetal liver. The present sequence is a peptide encoded by a single exon  
 CC nucleic acid probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 50 AA;  
 XX  
 Query Match 36.5%; Score 50; DB 22; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-43;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 26 RKKTFLSVHEVMAVENYAKDSLQWITDQYNKESDPDKYHFRIFRYLKVQRQ 75  
 Db 1 RKKTFLSVHEVMAVENYAKDSLQWITDQYNKESDPDKYHFRIFRYLKVQRQ 50  
 XX  
 RESULT 11  
 ABB19524  
 ID ABB19524 standard; Protein; 50 AA.  
 XX  
 AC ABB19524;  
 XX  
 DT 23-JAN-2002 (first entry)  
 XX  
 DE Protein #1523 encoded by probe for measuring heart cell gene expression.  
 XX  
 KW Human; gene expression; heart; microarray; vascular system;  
 KW cardiovascular disease; hypertension; cardiac arrhythmia;  
 KW congenital heart disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157274-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 DT 30-JAN-2001; 2001WO-US00666.  
 XX

PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-48899/53.  
 XX  
 PT Single exon nucleic acid probes for analyzing gene expression in human  
 PT hearts -  
 XX  
 PS Claim 15; SEQ ID NO 21294; 530bp; English.  
 XX  
 CC The present invention relates to single exon nucleic acid probes for  
 CC measuring human gene expression in a sample derived from human heart (see  
 CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
 CC probe. The probes may be used for predicting, measuring and displaying  
 CC gene expression in samples derived from the human heart via microarrays.  
 CC By measuring gene expression, the probes are useful for predicting,  
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
 CC human heart and vascular system e.g. cardiovascular disease,  
 CC hypertension, cardiac arrhythmias and congenital heart disease.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 50 AA;

Query Match 36.5%; Score 50; DB 22; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-43;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 RKKTFPSVHEVMAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKVRQ 75  
 DB 1 RKKTFPSVHEVMAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKVRQ 50

RESULT 12  
 AAM54871  
 ID AAM54871 standard; Protein; 50 AA.  
 XX  
 AC AAM54871;  
 XX  
 DT 05-NOV-2001 (first entry)  
 XX  
 DE Human brain expressed single exon probe encoded protein SEQ ID NO: 26976.  
 XX  
 KW Human; brain expressed exon; gene expression analysis; probe;  
 KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
 KW epilepsy; cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157275-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00667.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX

PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-483446/52.  
 DR  
 XX  
 PT Single exon nucleic acid probes for analyzing gene expression in human  
 PT brains -  
 XX  
 PS Example 4; SEQ ID NO: 26976; 650bp + Sequence listing; English.  
 XX  
 CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC brain. They can be used to measure gene expression in brain cell samples,  
 CC which may enable the diagnosis and improved treatment of nervous system  
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC epilepsy and cancers. The present sequence is a protein encoded by one of  
 CC the probes of the invention.  
 CC  
 XX  
 SQ Sequence 50 AA;

Query Match 36.5%; Score 50; DB 22; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-43;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 RKKTFPSVHEVMAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKVRQ 75  
 DB 1 RKKTFPSVHEVMAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKVRQ 50

RESULT 13  
 AAM67252  
 ID AAM67252 standard; Protein; 50 AA.  
 XX  
 AC AAM67252;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human bone marrow expressed probe encoded protein SEQ ID NO: 27558.  
 XX  
 KW Human; bone marrow expressed exon; gene expression analysis; probe;  
 KW microarray; cancer; leukemia; lymphoma; myeloma.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157276-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00668.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-48890/53.  
 DR  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human bone marrow -  
 XX  
 PS Example 4; SEQ ID NO: 27558; 650bp + Sequence listing; English.  
 XX  
 CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow

CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukemia and myeloma. The present sequence is a  
 CC protein encoded by one of the probes of the invention.

XX Sequence 50 AA;

Query Match 36.5%; Score 50; DB 22; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-43;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 RKKTFLSVHEWMAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKYQRQ 75  
 Db 1 RKKTFLSVHEWMAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKYQRQ 50

#### RESULT 14

AA015096  
 ID AAM15096 standard; Protein; 50 AA.

XX AAM15096;

XX 12-OCT-2001 (first entry)

XX Peptide #1530 encoded by probe for measuring cervical gene expression.

XX Probe; human; microarray; gene expression; cervical epithelial cell;

XX cervical cancer.

XX Homo sapiens.

XX WO200157278-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00670.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488901/53.

XX Claim 27; SEQ ID No 19922; 487bp; English.

XX The present invention relates to human single exon nucleic acid probes  
 CC (SENP; see A1110068-A128459). The present sequence is a peptide encoded  
 CC by one such probe. The SENPs are derived from human Hela cells. The SENPs  
 CC can be used to produce a single exon microarray, which can be used for  
 CC measuring human gene expression in a sample derived from human cervical  
 CC epithelial cells. By measuring gene expression, the probes are therefore  
 CC useful in grading and/or staging of diseases of the cervix, notably  
 CC cervical cancer.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 50 AA;

Query Match 36.5%; Score 50; DB 22; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-43;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 RKKTFLSVHEWMAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKYQRQ 75  
 Db 1 RKKTFLSVHEWMAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKYQRQ 50

RESULT 15  
 AAM27545  
 ID AAM27545 standard; Protein; 50 AA.

XX AAM27545;

XX 17-OCT-2001 (first entry)

XX Peptide #1582 encoded by probe for measuring placental gene expression.

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder.

XX Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-48897/53.

XX Claim 27; SEQ ID No 27814; 654bp; English.

XX The present invention relates to single exon nucleic acid probes (SENP;  
 CC see A113315-A157546). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for producing a microarray for  
 CC predicting, measuring and displaying gene expression in samples derived  
 CC from human placenta. The probes are useful for antenatal diagnosis of  
 CC human genetic disorders.

XX Sequence 50 AA;

Query Match 36.5%; Score 50; DB 22; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-43;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 RKKTFLSVHEWMAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKYQRQ 75  
 Db 1 RKKTFLSVHEWMAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKYQRQ 50

#### RESULT 16

AA02833  
 ID AAM02833 standard; Protein; 50 AA.

XX AAM02833;

XX 09-OCT-2001 (first entry)

XX Peptide #1515 encoded by probe for measuring breast gene expression.

KW Probe: human; breast disease; breast cancer; development disorder;  
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
 XX Homo sapiens.  
 OS  
 XX MO200157270-A2.  
 PN  
 XX  
 XX 09-AUG-2001.  
 PD  
 XX  
 XX 29-JAN-2001; 2001MO-US00661.  
 PF  
 XX 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2001-476286/51.  
 DR  
 XX Novel single exon nucleic acid probe used to measuring gene expression  
 PT in a human breast -  
 XX  
 XX Claim 27; SEQ ID NO 11573; 322pp; English.  
 PS  
 XX The present invention relates to novel single exon nucleic acid probes  
 CC (see AAI00010-AA110067). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for measuring human gene expression in  
 CC a human breast sample, where the probe hybridises at high stringency to a  
 CC nucleic acid expressed in the human breast. The probes are useful for  
 CC predicting, diagnosing, grading, staging, monitoring and prognosing  
 CC diseases of the human breast, particularly those diseases with polygenic  
 CC aetiology. The diseases include: breast cancer, disorders of development,  
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative  
 CC breast disease and non-carcinoma tumours.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 50 AA;  
 SQ  
 Query Match 36.5%; Score 50; DB 22; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-43;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 26 RKKTFPLSVHEVMAVENYAKDSLQWITDQYNKESDDKTHFRIFVLYKQVRQ 75  
 |||||||  
 Db 1 RKKTFPLSVHEVMAVENYAKDSLQWITDQYNKESDDKTHFRIFVLYKQVRQ 50  
 RESULT 17  
 ABG36903  
 ID ABG36903 standard; Peptide; 50 AA.  
 XX  
 AC ABG36903;  
 XX  
 XX 19-AUG-2002 (first entry)  
 DT  
 XX Human peptide encoded by genome-derived single exon probe SEQ ID 26568.  
 DB  
 XX Human; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karsagen syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;

KW hyaline membrane disease.  
 XX  
 XX Homo sapiens.  
 OS  
 XX MO200186003-A2.  
 PN  
 XX  
 XX 15-NOV-2001.  
 PD  
 XX  
 XX 30-JAN-2001; 2001MO-US00665.  
 PF  
 XX 04-FEB-2000; 2000US-180312P.  
 PR 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2002-114183/15.  
 DR  
 XX Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples -  
 XX  
 XX Claim 27; SEQ ID NO 26568; 634pp; English.  
 PS  
 XX The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridise at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryotic lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarray having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene  
 CC expression analysis, and for identifying exons in a gene, particularly  
 CC using human lung derived mRNA and for the study of lung diseases  
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
 CC pulmonary alveolar proteinosis, Karsagen syndrome, fibrocystic  
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
 CC and hyaline membrane disease. The present sequence is a peptide/protein  
 CC encoded by a single exon probe of the invention.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 50 AA;  
 SQ  
 Query Match 36.5%; Score 50; DB 23; Length 50;

Best Local Similarity 100.0%; Pred. No. 2,3e-43;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 RKRTFLSVHEVMAVENYAKSLQWITDOYNKSDKXHFRTLYKXQRO 75  
DB 1 RKRTFLSVHEVMAVENYAKSLQWITDOYNKSDKXHFRTLYKXQRO 50

## RESULT 18

AAU79863 standard; Peptide: 49 AA.

AAU79863;

15-JUL-2002 (first entry)

Human cystatin-8 (Zcys8) antigenic fragment #11.

Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;

KW spermatogenesis; seminal fluid viscosity; cryopreserved sperm;

KM sperm motility; fertilisation; antigenic peptide.

OS Homo sapiens.

WO200220567-A2.

14-MAR-2002.

29-AUG-2001, 2001WO-US26868.

01-SEP-2000; 2000US-230230P.

(ZYMO) ZYMOGENETICS INC.

Holloway JL, Gao Z, Bishop PD;

WPI; 2002-383044/41.

Novel isolated mammalian cystatin-8 polypeptide useful for promoting

spermatogenesis, and inhibiting cancer procoagulant protein which leads

to inhibition of thrombotic events associated with cancer -

Claim 2; Page 97-98; 100pp; English.

The invention describes an isolated mammalian cystatin-8 (Zcys8)

polypeptide (I). (I) is useful for: inhibiting cancer procoagulant

protein in an individual and thus inhibiting the thrombotic events

associated with cancer; promoting spermatogenesis; modulating seminal

fluid viscosity; enhancing viability of cryopreserved sperm; sperm

motility and fertilisation; and as antigenic peptides to generate

antibodies. Zcys8 is useful as research reagent for characterising sites

of interaction between Zcys8 and its receptor. Zcys8 is useful in

enhancing fertilisation during assisted reproduction in humans and in

animals. Anti-(I) antibodies are useful to screen biological samples like

blood, urine, saliva, tissue biopsy and autopsy material in vitro for the

presence of Zcys8. The antibodies are also useful to isolate large

quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.

The polynucleotide encoding (I) is useful to detect and to localise the

expression of a Zcys8 gene in a biological sample and Zcys8

oligonucleotide probes are useful for in vivo diagnosis. The

polynucleotide encoding (I) is useful in determining whether a subject's

chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene

copy number changes, insertions, deletions, restriction site changes and

rearrangements and genetic alterations that inactivate the Zcys8 gene.

This sequence represents an antigenic peptide of human cystatin-8

(Zcys8).

Sequence 49 AA;

Query Match 35.8%; Score 49; DB 23; Length 49;

Best Local Similarity 100.0%; Pred. No. 2,4e-42;

Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 DQYNKSDKXHFRTLYKXQROVTDHLEHYNVEMQWTTCKXPERTTN 100  
DB 1 DQYNKSDKXHFRTLYKXQROVTDHLEHYNVEMQWTTCKXPERTTN 49

## RESULT 19

AAU79867 standard; Peptide: 48 AA.

AAU79867;

15-JUL-2002 (first entry)

Human cystatin-8 (Zcys8) antigenic fragment #15.

Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;

KW spermatogenesis; seminal fluid viscosity; cryopreserved sperm;

KM sperm motility; fertilisation; antigenic peptide.

OS Homo sapiens.

WO200220567-A2.

14-MAR-2002.

29-AUG-2001, 2001WO-US26868.

01-SEP-2000; 2000US-230230P.

(ZYMO) ZYMOGENETICS INC.

Holloway JL, Gao Z, Bishop PD;

WPI; 2002-383044/41.

Novel isolated mammalian cystatin-8 polypeptide useful for promoting

spermatogenesis, and inhibiting cancer procoagulant protein which leads

to inhibition of thrombotic events associated with cancer -

Claim 2; Page 99; 100pp; English.

The invention describes an isolated mammalian cystatin-8 (Zcys8)

polypeptide (I). (I) is useful for: inhibiting cancer procoagulant

protein in an individual and thus inhibiting the thrombotic events

associated with cancer; promoting spermatogenesis; modulating seminal

fluid viscosity; enhancing viability of cryopreserved sperm; sperm

motility and fertilisation; and as antigenic peptides to generate

antibodies. Zcys8 is useful as research reagent for characterising sites

of interaction between Zcys8 and its receptor. Zcys8 is useful in

enhancing fertilisation during assisted reproduction in humans and in

animals. Anti-(I) antibodies are useful to screen biological samples like

blood, urine, saliva, tissue biopsy and autopsy material in vitro for the

presence of Zcys8. The antibodies are also useful to isolate large

quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.

The polynucleotide encoding (I) is useful to detect and to localise the

expression of a Zcys8 gene in a biological sample and Zcys8

oligonucleotide probes are useful for in vivo diagnosis. The

polynucleotide encoding (I) is useful in determining whether a subject's

chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene

copy number changes, insertions, deletions, restriction site changes and

rearrangements and genetic alterations that inactivate the Zcys8 gene.

This sequence represents an antigenic peptide of human cystatin-8

(Zcys8).

Sequence 48 AA;

Query Match 35.0%; Score 48; DB 23; Length 48;

Best Local Similarity 100.0%; Pred. No. 2,4e-41;

Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 NVEMQWTTCKXPERTTNVCPQERELHKOVNCFPSVFAVPMWFOYKILNK 132  
DB 1 NVEMQWTTCKXPERTTNVCPQERELHKOVNCFPSVFAVPMWFOYKILNK 48

**RESULT 20**

AAU79860  
ID AAU79860 standard; Peptide; 46 AA.  
....

AC AAU79860;

DT 15-JUL-2002 (first entry)

DE Human cystatin-8 (Zcys8) antigenic fragment #8.

Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;

KM spermatogenesis; seminal fluid viscosity; cryopreserved sperm;  
KW sperm motility; fertilisation; antigenic peptide.

**KW** sperm motility; fertilisation; antigenic peptide.

**Homo sapiens**

PN WO200220567-A2.

PD 14-MAR-2002.

PF 29-AUG-2001; 2001WO-US26868.

PR 01-SEP-2000; 2000US-230230P.

PA (ZYMO ) ZYMOGENETICS INC.

PI Holloway JL, Gao Z, Bishop PD;

DR WPI; 2002-383044/41.

PT Novel isolated mammalian cystatin-8 polypeptide useful for promoting spermatogenesis, and inhibiting cancer procoagulant protein which leads to inhibition of thrombotic events associated with cancer -

PS Claim 2; Page 97; 100pp; English.

CC The invention describes an isolated mammalian cystatin-8 (Zcys8)

CC protein in an individual and thus inhibiting the thrombotic events

Query Match	33.6%	Score 46;	DB 23;	Length 46;
Best Local Similarity	100.0%	Pred. NO. 2.6e-39;		
Matches 46; Conservative	0;	Mismatches	0;	Gaps 0

QY	Db
55	1
56	1
57	1
58	1
59	1
60	1
61	1
62	1
63	1
64	1
65	1
66	1
67	1
68	1
69	1
70	1
71	1
72	1
73	1
74	1
75	1
76	1
77	1
78	1
79	1
80	1
81	1
82	1
83	1
84	1
85	1
86	1
87	1
88	1
89	1
90	1
91	1
92	1
93	1
94	1
95	1
96	1
97	1
98	1
99	1
100	1

## RESULT 21

AAU79856  
ID AU79856 standard; Peptide; 36 AA

AC AAU79856;

DT 15-JUL-2002 (first entry)

DB Human cystatin-8 (Zcys8) antigenic fragment #4

KW Cystatin-B; Zcys8; cancer; procoagulant protein; thrombosis;

KW spermatogenesis; seminal fluid viscosity; cryopreserved sperm; sperm motility; fertilisation; antigenic fragment.

KW sperm motility; fertilisation; antigenic fragment.

**Homo sapiens.**

PN W0200220567-A2

PD 14-MAR-2002.

PF 29-AUG-2001; 2001WO-US26868.

PR 01-SEP-2000; 2000US-230230P.

PA (ZYMO) ZYMOGENETICS INC.

PI Holloway JL, Gao Z, Bishop PD;

DR WPI; 2002-383044/41.

PT Novel isolated mammalian cystatin-8 polypeptide useful for promoting spermatogenesis, and inhibiting cancer procoagulant protein which leads to inhibition of thrombotic events associated with cancer -

PS Claim 2; Page 95; 100pp; English.

The invention describes an isolated mammalian cystatin-8 (Zcys8) polypeptide (I). (I) is useful for: inhibiting cancer procoagulant protein in an individual and thus inhibiting the thrombotic events associated with cancer; promoting spermatogenesis, modulating seminal fluid viscosity, enhancing viability of cryopreserved sperm, sperm motility and fertilisation; and as antigenic peptides to generate antibodies. Zcys8 is useful as research reagent for characterising sites of interaction between Zcys8 and its receptor. Zcys8 is useful in enhancing fertilisation during assisted reproduction in humans and in animals. Anti-(I) antibodies are useful to screen biological samples like blood, urine, saliva, tissue biopsy and autopsy material *in vitro* for the presence of Zcys8. The antibodies are also useful to isolate large quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes. The polynucleotide encoding (I) is useful to detect and to localise the expression of a Zcys8 gene in a biological sample and Zcys8 oligonucleotide probes are useful for *in vivo* diagnosis. The polynucleotide encoding (I) is useful in determining whether a subject's chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene copy number changes, insertions, deletions, restriction site changes and rearrangements and genetic alterations that inactivate the Zcys8 gene. This sequence represents an antigenic fragment of human cystatin-8 (Zcys8).

**SQ Sequence 36 AA;**

Query Match	26.3%;	Score 36;	DB 23;	Length 36;
Best Local Similarity	100.0%;	Pred. No. 3.4e-29;		
Matches 36;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

[illegible]

RESULT 22	
AAU79858	
ID	AAU79858 standard; Peptide; 35 AA
XX	



KW Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;  
 KW spermatogenesis; seminal fluid viscosity; cryopreserved sperm;  
 KW sperm motility; fertilisation; antigenic peptide.  
 OS Homo sapiens.  
 XX  
 XX WO200220567-A2.  
 PN  
 XX 14-MAR-2002.  
 PD  
 XX 29-AUG-2001; 2001WO-US26668.  
 PF  
 XX 01-SEP-2000; 2000US-230230P.  
 PR  
 XX (ZYMO ) ZYMOGENETICS INC.  
 PA  
 XX Holloway JL, Gao Z, Bishop PD;  
 PI WPI; 2002-383044/41.  
 DR  
 XX Novel isolated mammalian cystatin-8 polypeptide useful for promoting  
 PT spermatogenesis, and inhibiting cancer procoagulant protein which leads  
 PT to inhibition of thrombotic events associated with cancer -  
 PS  
 XX Claim 2; Page 96; 100pp; English.  
 PS  
 XX The invention describes an isolated mammalian cystatin-8 (Zcys8)  
 CC polypeptide (I). (I) is useful for: inhibiting cancer procoagulant  
 CC protein in an individual and thus inhibiting the thrombotic events  
 CC associated with cancer; promoting spermatogenesis, modulating seminal  
 CC fluid viscosity, enhancing viability of cryopreserved sperm, sperm  
 CC motility and fertilisation; and as antigenic peptides to generate  
 CC antibodies. Zcys8 is useful as research reagent for characterising sites  
 CC of interaction between Zcys8 and its receptor. Zcys8 is useful in  
 CC enhancing fertilisation during assisted reproduction in humans and in  
 CC animals. Anti-(I) antibodies are useful to screen biological samples like  
 CC blood, urine, saliva, tissue biopsy and autopsy material in vitro for the  
 CC presence of Zcys8. The antibodies are also useful to isolate large  
 CC quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.  
 CC The polynucleotide encoding (I) is useful to detect and to localise the  
 CC expression of a Zcys8 gene in a biological sample and Zcys8  
 CC oligonucleotide probes are useful for in vivo diagnosis. The  
 CC polynucleotide encoding (I) is useful in determining whether a subject's  
 CC chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene  
 CC copy number changes, insertions, deletions, restriction site changes and  
 CC rearrangements and genetic alterations that inactivate the Zcys8 gene.  
 CC This sequence represents an antigenic peptide of human cystatin-8  
 CC (Zcys8).  
 CC  
 XX  
 SQ Sequence 27 AA;  
 Query Match 19.7%; Score 27; DB 23; Length 27;  
 Best Local Similarity 100.0%; Pred. No. 4e-20;  
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 52 DQYKESDDKXHFRIPLVKVQRQVTD 78  
 Db 1 DQYKESDDKXHFRIPLVKVQRQVTD 27  
 RESULT 25  
 AAU79861  
 ID AAU79861 standard; Peptide; 24 AA.  
 AC AAU79861;  
 XX  
 XX 15-JUL-2002 (first entry)  
 DE Human cystatin-8 (Zcys8) antigenic fragment #9.  
 KW Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;  
 KW spermatogenesis; seminal fluid viscosity; cryopreserved sperm;  
 KW sperm motility; fertilisation; antigenic peptide.

XX  
 OS Homo sapiens.  
 XX  
 XX WO200220567-A2.  
 PN  
 XX 14-MAR-2002.  
 PD  
 XX 29-AUG-2001; 2001WO-US26668.  
 PF  
 XX 01-SEP-2000; 2000US-230230P.  
 PR  
 XX (ZYMO ) ZYMOGENETICS INC.  
 PA  
 XX Holloway JL, Gao Z, Bishop PD;  
 PI WPI; 2002-383044/41.  
 DR  
 XX Novel isolated mammalian cystatin-8 polypeptide useful for promoting  
 PT spermatogenesis, and inhibiting cancer procoagulant protein which leads  
 PT to inhibition of thrombotic events associated with cancer -  
 PS  
 XX Claim 2; Page 97; 100pp; English.  
 PS  
 XX The invention describes an isolated mammalian cystatin-8 (Zcys8)  
 CC polypeptide (I). (I) is useful for: inhibiting cancer procoagulant  
 CC protein in an individual and thus inhibiting the thrombotic events  
 CC associated with cancer; promoting spermatogenesis, modulating seminal  
 CC fluid viscosity, enhancing viability of cryopreserved sperm, sperm  
 CC motility and fertilisation; and as antigenic peptides to generate  
 CC antibodies. Zcys8 is useful as research reagent for characterising sites  
 CC of interaction between Zcys8 and its receptor. Zcys8 is useful in  
 CC enhancing fertilisation during assisted reproduction in humans and in  
 CC animals. Anti-(I) antibodies are useful to screen biological samples like  
 CC blood, urine, saliva, tissue biopsy and autopsy material in vitro for the  
 CC presence of Zcys8. The antibodies are also useful to isolate large  
 CC quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.  
 CC The polynucleotide encoding (I) is useful to detect and to localise the  
 CC expression of a Zcys8 gene in a biological sample and Zcys8  
 CC oligonucleotide probes are useful for in vivo diagnosis. The  
 CC polynucleotide encoding (I) is useful in determining whether a subject's  
 CC chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene  
 CC copy number changes, insertions, deletions, restriction site changes and  
 CC rearrangements and genetic alterations that inactivate the Zcys8 gene.  
 CC This sequence represents an antigenic peptide of human cystatin-8  
 CC (Zcys8).  
 CC  
 XX  
 SQ Sequence 24 AA;  
 Query Match 17.5%; Score 24; DB 23; Length 24;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-17;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 74 ROYTDHLEHYLANVEMQWTCOKPE 97  
 Db 1 ROYTDHLEHYLANVEMQWTCOKPE 24  
 RESULT 26  
 AAU79855  
 ID AAU79855 standard; Peptide; 22 AA.  
 AC AAU79855;  
 XX  
 XX 15-JUL-2002 (first entry)  
 DE Human cystatin-8 (Zcys8) antigenic fragment #3.  
 KW Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;  
 KW spermatogenesis; seminal fluid viscosity; cryopreserved sperm;  
 KW sperm motility; fertilisation; antigenic fragment.  
 OS Homo sapiens.



```

PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PMD, Myers EW,
XX
XX WPI, 2001-656860/75.
XX
XX N-PSDB; ABL03466.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure, SEQ ID NO 4881; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB16176-AB170511), expressed DNA
XX sequences (AB101840-AB116175) and the encoded proteins
XX (AB57737-AB572072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1252 AA;
SQ
Query Match 5.8%; Score 8; DB 22; Length 1252;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 LLLAIIILT 17
11 LLLAIIILT 18
Db 11 LLLAIIILT 18

RESULT 29
AAU66887
ID AAU66887 standard; Protein; 60 AA.
XX
XX AAU66887;
XX
XX 13-FEB-2002 (first entry)
XX
XX Propionibacterium acnes immunogenic protein #2783.
XX
XX SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteopathic; neuroprotectant.
XX
XX Propionibacterium acnes.
XX
XX WO200181581-A2.
XX
XX 01-NOV-2001.
XX
XX 20-APR-2001; 2001WO-US12865.
XX
XX 21-APR-2000; 2000US-199047P.
XX
XX 02-JUN-2000; 2000US-208841P.
XX
XX 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhactia A;
XX
XX L'malsomeuve J, Zhang Y, Jen S, Carter D;
XX
XX WPI; 2001-616774/71.
XX
XX N-PSDB; AASS9763.
XX

```

```

PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
XX Example 1; SEQ ID No 28082; 1069pp; English.
XX
XX Sequences AAU9105-AAU68017 represent Propionibacterium acnes immunogenic
XX polypeptides. The proteins and their associated DNA sequences are used in
XX the treatment, prevention and diagnosis of medical conditions caused by
XX P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
XX P. acnes is also involved in infections of bone, joints and the central
XX nervous system, however it is particularly involved in the inflammatory
XX lesions associated with acne vulgaris. A method for detecting the
XX presence or absence of P. acnes in a patient comprises contacting a
XX sample with a binding agent that binds to the proteins of the invention
XX and determining the amount of bound protein in the sample. The
XX polypeptides may be used as antigens in the production of antibodies
XX specific for P. acnes proteins. These antibodies can be used to
XX downregulate expression and activity of P. acnes polypeptides and
XX therefore treat P. acnes infections. The antibodies may also be used as
XX diagnostic agents for determining P. acnes presence, for example, by
XX enzyme linked immunosorbent assay (ELISA).
XX
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 60 AA;
SQ
Query Match 5.1%; Score 7; DB 22; Length 60;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 106 RELHKQV 112
49 RELHKQV 55
Db 49 RELHKQV 55

RESULT 30
ABG03232
ID ABG03232 standard; Protein; 97 AA.
XX
XX ABG03232;
XX
XX 13-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #3223.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX N-PSDB; AAS67419.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX

```

XX Claim 20; SEQ ID No 33591; 103bp; English.  
PS  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 97 AA;  
Query Match 5.1%; Score 7; DB 22; Length 97;  
Best Local Similarity 100.0%; Pred. No. 32;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 10 LLLALL 16  
Db 27 LLLALL 33  
RESULT 31  
AAG21482  
ID AAG21482 standard; Protein; 98 AA.  
XX  
AC AAG21482;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 24053.  
XX  
KM Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
PD  
XX 06-SEP-2000.  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 99US-0131449.

PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135253.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.

```

PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.

```

```

PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      5.1%; Score 7; DB 21; Length 98;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      10 L1L1L1L 16
Db      83 L1L1L1L 89

RESULT 32
AAG48767
ID AAG48767 standard; Protein; 98 AA.
XX
AC AAG48767;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61618.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX
PD 06-SEP-2000.
XX
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 29-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 16-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 04-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.

```

PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139486.  
PR 18-JUN-1999; 99US-0139489.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140699.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 19-JUL-1999; 99US-0144337.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144633.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145087.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156586.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159684.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.

PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
  
Query Match 5.1%; Score 7; DB 21; Length 98;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Oy 10 LLAALL 16  
Db 83 LLAALL 89  
  
RESULT 33  
ID AAG21481 standard; Protein: 100 AA.  
AC AAG21481;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 24052.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 09-MAR-1999; 99US-0123180.  
PR 23-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 29-MAR-1999; 99US-0126264.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134321.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 21-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 22-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 28-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149922.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.  
Query Match 5.1%; Score 7; DB 21; Length 100;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 10 LLLALL 16  
Db 85 LLLALL 91  
RESULT 34  
ID AAC48766 standard; Protein; 100 AA.  
AC AAC48766;  
XX 18-OCT-2000 (first entry)  
DT Arabidopsis thaliana protein fragment SEQ ID NO: 61617.  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 61617.  
XX Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
OS Arabidopsis thaliana.  
XX EP1033405-A2.  
XX 06-SEP-2000.  
PF 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123568.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140981.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142390.  
PR 09-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142970.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147716.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152263.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 15-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 14-OCT-1999; 99US-0159684.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 25-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match

5.1%; Score 7; DB 21; Length 100;

Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLALL 16  
|||  
|||  
Db 85 LLLALL 91

## RESULT 35

AA642196  
ID AA642196 standard; Protein, 102 AA.

XX AA642196;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 52593.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX Arabidopsis thaliana.

PN EPI033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 23-APR-1999; 99US-0130891.

XX 28-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 30-APR-1999; 99US-0132407.

XX 04-MAY-1999; 99US-0132485.

XX 05-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 06-MAY-1999; 99US-0132487.

XX 07-MAY-1999; 99US-0132863.

XX 11-MAY-1999; 99US-0134256.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.

XX 14-MAY-1999; 99US-0134370.

XX 18-MAY-1999; 99US-0134768.

XX 19-MAY-1999; 99US-0134941.

XX 20-MAY-1999; 99US-0135124.

XX 21-MAY-1999; 99US-0135353.

XX 24-MAY-1999; 99US-0135629.

XX 25-MAY-1999; 99US-0136021.

XX 27-MAY-1999; 99US-0136392.

XX 28-MAY-1999; 99US-0136782.

XX 01-JUN-1999; 99US-0137222.

XX 03-JUN-1999; 99US-0137528.

XX 04-JUN-1999; 99US-0137502.

XX 07-JUN-1999; 99US-0137724.

XX 08-JUN-1999; 99US-0138094.

XX 10-JUN-1999; 99US-0138540.

XX 10-JUN-1999; 99US-0138847.

XX 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139753.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140921.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147018.

PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147320.

PR 05-AUG-1999; 99US-0147360.

PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.

PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.

```

PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 5.1%; Score 7; DB 21; Length 102;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 10 LLLALL 16
DB 85 LLLALL 91

```

```

RESULT 36
AAU83235
ID AAU83235 standard; Protein; 112 AA.
XX
XX AAU83235;
AC
XX 08-MAY-2002 (first entry)
DT
XX
XX Novel secreted protein Z935205G2P.
DE
XX
XX Protein secretion; mammalian secreted polypeptide; MSP.
KM
XX Homo sapiens.
OS
XX WO200202621-A2.
PN
XX 10-JAN-2002.
PD
XX
XX 28-JUN-2001; 2001MO-US20638.
PF
XX
XX 30-JUN-2000; 2000US-215446P.
PR
XX (ZYMO ) ZYMOGENETICS INC.
PA
XX Sheppard PO, Presnell SR;
PI
XX
XX WPI; 2002-147999/19.
DR
XX N-PSDB; ABK33150.
DR
XX
XX Novel isolated mammalian secreted polypeptide useful in therapeutic and
PT diagnostic methods, to direct secretion of other proteins of interest
PT from host cell, as educational tools, and as laboratory practicum kits
PT
XX
XX Claim 12; Page 389; 397pp; English.
PS
XX
XX The invention describes an isolated mammalian secreted polypeptide (MSP)
CC (1). (1) is useful to direct the secretion of other proteins of interest
CC from a host cell, to monitor secretion of proteins, to degenerate
CC sequences comprising all nucleotide sequences encoding a particular
CC polypeptide, to screen for cell metabolism effecting receptors, for
CC identifying new target receptors and drug design, for identifying, for
CC protein purification, for determining the weight of expressed MSP
CC polypeptides as a ratio to total protein expressed, for identifying
CC peptide cleavage sites, for coupling amino and carboxy terminal tags, for
CC amino acid sequence analysis, for monitoring biological activities of the
CC protein in vitro and in vivo, and to teach analytical skills and as
CC reagents for the study of cells, receptors, and other binding molecules.
CC The polynucleotide is useful for radiation hybrid mapping, and somatic
CC cell genetic technique developed for constructing high-resolution,
CC contiguous maps of mammalian chromosomes. Reagents disclosed in the
CC invention may be used to detect metabolic abnormalities characterised by
CC over or under production of the protein. This is the amino acid sequence
CC of a mammalian secreted polypeptide, described in the method of the
CC invention.
CC
XX
XX
SQ Sequence 112 AA;

```

```

Query Match 5.1%; Score 7; DB 23; Length 112;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 7 ALQLLA 13
DB 12 ALQLLA 18

```

```

RESULT 37
AAG42195
ID AAG42195 standard; Protein; 128 AA.
XX
XX AAG42195;
AC

```

XX 18-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 52592.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX Arabidopsis thaliana.  
XX EP1033405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
XX 09-MAR-1999; 99US-0123548.  
XX 23-MAR-1999; 99US-0125788.  
XX 25-MAR-1999; 99US-0126264.  
XX 29-MAR-1999; 99US-0126785.  
XX 01-APR-1999; 99US-0127462.  
XX 06-APR-1999; 99US-0128234.  
XX 08-APR-1999; 99US-0128714.  
XX 16-APR-1999; 99US-0129845.  
XX 19-APR-1999; 99US-0130077.  
XX 21-APR-1999; 99US-0130049.  
XX 23-APR-1999; 99US-0130510.  
XX 28-APR-1999; 99US-0130891.  
XX 30-APR-1999; 99US-0131449.  
XX 30-APR-1999; 99US-0132048.  
XX 30-APR-1999; 99US-0132407.  
XX 04-MAY-1999; 99US-0132484.  
XX 05-MAY-1999; 99US-0132485.  
XX 06-MAY-1999; 99US-0132486.  
XX 07-MAY-1999; 99US-0132487.  
XX 11-MAY-1999; 99US-0132863.  
XX 11-MAY-1999; 99US-0134256.  
XX 14-MAY-1999; 99US-0134218.  
XX 14-MAY-1999; 99US-0134219.  
XX 14-MAY-1999; 99US-0134221.  
XX 14-MAY-1999; 99US-0134370.  
XX 18-MAY-1999; 99US-0134768.  
XX 19-MAY-1999; 99US-0134941.  
XX 20-MAY-1999; 99US-0135124.  
XX 21-MAY-1999; 99US-0135353.  
XX 24-MAY-1999; 99US-0135629.  
XX 25-MAY-1999; 99US-0136021.  
XX 27-MAY-1999; 99US-0136392.  
XX 28-MAY-1999; 99US-0136782.  
XX 01-JUN-1999; 99US-0137222.  
XX 03-JUN-1999; 99US-0137528.  
XX 04-JUN-1999; 99US-0137502.  
XX 07-JUN-1999; 99US-0137724.  
XX 08-JUN-1999; 99US-0138094.  
XX 10-JUN-1999; 99US-0138540.  
XX 10-JUN-1999; 99US-0138847.  
XX 14-JUN-1999; 99US-0139119.  
XX 16-JUN-1999; 99US-0139452.  
XX 16-JUN-1999; 99US-0139453.  
XX 17-JUN-1999; 99US-0139452.  
XX 18-JUN-1999; 99US-0139454.  
XX 18-JUN-1999; 99US-0139456.  
XX 18-JUN-1999; 99US-0139457.  
XX 18-JUN-1999; 99US-0139458.  
XX 18-JUN-1999; 99US-0139459.  
XX 18-JUN-1999; 99US-0139460.  
XX 18-JUN-1999; 99US-0139461.  
XX 18-JUN-1999; 99US-0139462.  
XX 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 27-JUL-1999; 99US-0145951.  
PR 28-JUL-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147319.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 06-AUG-1999; 99US-0147433.  
PR 09-AUG-1999; 99US-0147935.  
PR 09-AUG-1999; 99US-0148171.  
PR 10-AUG-1999; 99US-0148319.  
PR 11-AUG-1999; 99US-0148341.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148655.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0150566.  
PR 25-AUG-1999; 99US-0150884.  
PR 26-AUG-1999; 99US-0150884.

PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0156599.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159283.  
 PR 13-OCT-1999; 99US-0159284.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 5.1%; Score 7; DB 21; Length 126;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 LLLALLL 16  
 Db 111 LLLALLL 117

RESULT 38  
 ABB89858  
 ID ABB89858 standard; Protein; 182 AA.  
 AC ABB89858;  
 XX  
 XX  
 DT 24-MAY-2002 (first entry)  
 XX  
 DB Human polypeptide SEQ ID NO 2234.  
 XX  
 KW Cytosolic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KM antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;

AC ABB89858;  
 XX  
 XX  
 DT 24-MAY-2002 (first entry)  
 XX  
 DB Human polypeptide SEQ ID NO 2234.  
 XX  
 KW Cytosolic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 KM antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;

KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; human; secreted protein.  
 OS Homo sapiens.  
 XX  
 XX WO200190304-A2.  
 PN  
 XX  
 XX 29-NOV-2001.  
 PD  
 XX  
 XX 18-MAY-2001; 2001MO-US16450.  
 PF  
 XX 19-MAY-2000; 2000US-205515P.  
 PR  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX Birse CR, Rosen CA;  
 PI Birse CR, Rosen CA;  
 XX WPI; 2002-122018/16.  
 DR N-PSDB; ABL90267.  
 DR  
 XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
 PT prevention of neural, immune system, muscular, reproductive,  
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
 PT disorders -  
 XX  
 XX Claim 11; SEQ ID NO 2234; 2081pp + Sequence listing; English.  
 PS  
 XX The invention relates to novel genes (ABL9449-ABL90853) and proteins  
 CC (ABB9040-ABB9044) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 CC hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pat\_sequences.  
 CC  
 XX  
 SO Sequence 182 AA;

Query Match 5.1%; Score 7; DB 23; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 LALLLTL 18  
 Db 88 LALLLTL 94

RESULT 39  
 ABB90713  
 ID ABB90713 standard; Protein; 184 AA.  
 AC ABB90713;  
 XX  
 XX  
 DT 29-JUL-2002 (first entry)  
 XX  
 XX  
 DE Chlamydia pneumoniae cp6668 protein, SEQ ID NO:375.  
 XX  
 XX Chlamydial infection; antigen; immunogen; vaccine; diagnosis;  
 KW human respiratory disease; cardiovascular disease; atherosclerosis;  
 KW coronary artery disease; carotid artery stenosis; myocardial infarction;  
 KW cerebrovascular disease; aortic aneurysm; claudication; stroke;  
 KM strain CML029.

AC ABB90713;  
 XX  
 XX  
 DT 29-JUL-2002 (first entry)  
 XX  
 XX  
 DE Chlamydia pneumoniae cp6668 protein, SEQ ID NO:375.  
 XX  
 XX Chlamydial infection; antigen; immunogen; vaccine; diagnosis;  
 KW human respiratory disease; cardiovascular disease; atherosclerosis;  
 KW coronary artery disease; carotid artery stenosis; myocardial infarction;  
 KW cerebrovascular disease; aortic aneurysm; claudication; stroke;  
 KM strain CML029.

```

XX OS Chlamydia pneumoniae.
XX PN WO200202606-A2.
XX PD 10-JAN-2002.
XX PE 03-JUL-2001; 2001WO-1B01445.
XX PR 03-JUL-2000; 2000GB-0016363.
XX PR 11-JUL-2000; 2000GB-0017047.
XX PR 21-JUL-2000; 2000GB-0017983.
XX PR 07-AUG-2000; 2000GB-0019368.
XX PR 18-AUG-2000; 2000GB-0020440.
XX PR 14-SEP-2000; 2000GB-0022583.
XX PR 10-NOV-2000; 2000GB-0027549.
XX PR 22-DEC-2000; 2000GB-0031706.
XX PA (CHIR-) CHIRON SPA.
XX PI Racti G, Grandi G;
XX PT WPI; 2002-154726/20.
XX DR N-PSDB; ABL91371.
XX PT Novel Chlamydia pneumoniae protein useful in the manufacture of a
XX PT medicament for treatment or prevention of infection due to Chlamydia,
XX PT preferably Chlamydia pneumoniae, and for diagnostic purposes
XX PS Claim 1; Page 185; 364pp; English.
XX CC Sequences ABB90526-ABB90715 represent novel proteins from Chlamydia
XX CC pneumoniae (strain CWL029), and ABL91184-ABL91373 represent DNA encoding
XX CC them. The proteins are predicted to be immunogenic and may therefore be
XX CC useful in vaccine production and for diagnostic purposes. Chlamydia
XX CC pneumoniae is a common cause of respiratory disease in humans, and is
XX CC also involved in the development of cardiovascular diseases such as
XX CC atherosclerosis, coronary artery disease, carotid artery stenosis,
XX CC myocardial infarction, cerebrovascular disease, aortic aneurysm,
XX CC claudication and stroke. The proteins and nucleic acids of the invention
XX CC may be used in vaccines and pharmaceutical compositions for the
XX CC prevention or treatment of chlamydial infections, particularly Chlamydia
XX CC pneumoniae infections. The proteins may also be used in the detection of
XX CC Chlamydia pneumoniae, and the nucleic acids may be used in PCR, branched
XX CC DNA probe assay or blotting techniques for determining Chlamydia
XX CC pneumoniae gene expression. The present sequence represents a
XX CC specifically claimed Chlamydia pneumoniae protein of the invention.
XX SQ Sequence 184 AA;
XX
XX Query Match 5.1%; Score 7; DB 23; Length 184;
XX Best Local Similarity 100.0%; Pred. No. 57;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 RKKTFLS 32
XX |||||
XX 26 RKKTFLS 32
Db
XX
XX RESULT 40
XX ABB68742
XX ID ABB68742 standard; Protein; 228 AA.
XX AC ABB68742;
XX DT 26-MAR-2002 (first entry)
XX XX Drosophila melanogaster polypeptide SEQ ID NO 33018.
XX DE Drosophila melanogaster polypeptide SEQ ID NO 33018.
XX XX Drosophila; developmental biology; cell signalling; insecticide;
XX KM pharmaceutical.
XX OS Drosophila melanogaster.

```

```

XX XX WO200171042-A2.
XX XX 27-SEP-2001.
XX PD 23-MAR-2001; 2001WO-US09231.
XX PE 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX XX (PEKE ) PE CORP NY.
XX PA Venter JC, Adams M, Li PWD, Myers EW;
XX PI WPI; 2001-656860/75.
XX DR N-PSDB; ABL12845.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions -
XX PS Disclosure; SEQ ID NO 33018; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins
XX CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 228 AA;
XX
XX Query Match 5.1%; Score 7; DB 22; Length 228;
XX Best Local Similarity 100.0%; Pred. No. 68;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 83 HLNVMQ 89
XX |||||
XX 83 HLNVMQ 89
Db
XX
XX RESULT 41
XX ABB70167
XX ID ABB70167 standard; Protein; 267 AA.
XX AC ABB70167;
XX DT 26-MAR-2002 (first entry)
XX XX Drosophila melanogaster polypeptide SEQ ID NO 37293.
XX DE Drosophila melanogaster polypeptide SEQ ID NO 37293.
XX XX Drosophila; developmental biology; cell signalling; insecticide;
XX KM pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX XX 23-MAR-2001; 2001WO-US09231.
XX PE 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX XX (PEKE ) PE CORP NY.
XX PA Venter JC, Adams M, Li PWD, Myers EW;
XX PI

```

XX WPI, 2001-65860/75.  
 DR N-PSDB; ABL14270.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from *Drosophila* and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Disclosure; SEQ ID NO 37293; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL1840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 267 AA;  
 SQ  
 Query Match 5.1%; Score 7; DB 22; Length 267;  
 Best Local Similarity 100.0%; Pred. No. 78;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 94 ||||| 100  
 DB 22 OKPETHN 28  
 RESULT 42  
 ABB53808  
 ID ABB53808 standard; Protein; 288 AA.  
 XX  
 AC ABB53808;  
 XX  
 DT 16-MAY-2002 (first entry)  
 XX  
 DE Lactococcus lactis protein yelI.  
 XX  
 KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.  
 XX  
 OS Lactococcus lactis IL1403.  
 XX  
 PN FR807446-A1.  
 XX  
 PD 12-OCT-2001.  
 XX  
 PF 11-APR-2000; 2000PR-0004630.  
 XX  
 PR 11-APR-2000; 2000PR-0004630.  
 XX  
 PA (INRG) INRA INST NAT RECH AGRONOMIQUE.  
 XX  
 PI Bolocline A, Sorokine A, Renault P, Ehrlich SD;  
 XX  
 DR WPI, 2002-043418/06.  
 XX  
 PT New nucleotide sequence useful in the identification or Lactococcus  
 PT lactis and related species -  
 XX  
 PS Claim 6; SEQ ID NO 510; 2504pp; French.  
 XX  
 CC The present invention is related to a Lactococcus lactis nucleotide  
 CC sequence (ABA90521) and related proteins (ABB53300-ABB5621). The  
 CC nucleic acid sequence is useful in the detection and/or amplification  
 CC of nucleic acid sequence, particularly to identify Lactococcus lactis or  
 CC related species. The proteins of the invention are useful for the  
 CC biosynthesis or biodegradation of a composition of interest. The  
 CC invention helps research in lactic bacteria, particularly useful in the

CC production of yogurt and cheese.  
 CC Note: The sequence data for this patent is based on equivalent patent  
 CC WO200177334 (published 18-OCT-2001) which is available in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX Sequence 288 AA;  
 SQ  
 Query Match 5.1%; Score 7; DB 23; Length 288;  
 Best Local Similarity 100.0%; Pred. No. 84;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 10 LLLATLL 16  
 DB 21 LLLATLL 27  
 RESULT 43  
 AAB15904  
 ID AAB15904 standard; Protein; 330 AA.  
 XX  
 AC AAB15904;  
 XX  
 DT 05-OCT-2000 (first entry)  
 XX  
 DE E. coli proliferation associated protein sequence SEQ ID NO:261.  
 XX  
 KW Escherichia coli; E. coli; proliferation; inhibition; screening;  
 KW antimicrobial; bacterial growth; antisense therapy; antibacterial.  
 XX  
 OS Escherichia coli.  
 XX  
 PN WO200044906-A2.  
 XX  
 PD 03-AUG-2000.  
 XX  
 PF 27-JAN-2000; 2000WO-US022200.  
 XX  
 PR 27-JAN-1999; 99US-0117405.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX  
 DR WPI, 2000-514822/46.  
 DR N-PSDB; AAB65909.  
 XX  
 PT Novel polynucleotides and polypeptides associated with microorganism  
 PT proliferation, used to identify inhibitors of bacterial growth and  
 PT proliferation, for use in antisense therapy -  
 XX  
 PS Claim 11; Page 190-191; 316pp; English.  
 XX  
 CC AAB65809 to AAB65889 and AAB66058 to AAB66138 represent nucleotide  
 CC sequences derived from *Escherichia coli* which inhibit E. coli  
 CC proliferation. AAB65890 to AAB66055 and AAB15886 to AAB16040 represent  
 CC nucleotide and protein sequences associated with E. coli proliferation.  
 CC AAB66056 and AAB66057 represent primers used for sequencing E. coli  
 CC proliferation inhibiting nucleotide inserts in an example from the  
 CC present invention. Methods from the present invention can be used to  
 CC identify a proliferation-regulated gene in a microorganism, by contacting  
 CC a microorganism with a proliferation-regulated gene activity inhibitory  
 CC nucleic acid identified in another organism, and determining if  
 CC inhibition occurs in the second microorganism. The nucleic acid sequences  
 CC identified as being required for bacterial growth and proliferation, can  
 CC be used for antisense therapy for killing bacteria.  
 XX  
 SQ Sequence 330 AA;  
 Query Match 5.1%; Score 7; DB 21; Length 330;  
 Best Local Similarity 100.0%; Pred. No. 95;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LALLTL 18  
 |||||  
 DB 88 LALLTL 94

RESULT 44  
 ID AAO22146 standard; Protein; 333 AA.  
 XX AAO22146;  
 AC AAO22146;  
 DT 03-OCT-2002 (first entry)  
 XX  
 DE Ramoplanin biosynthetic ORF 1 protein.  
 XX  
 KM Ramoplanin; ramoplanin biosynthetic pathway polypeptide; antibiotic;  
 KM biosynthesis gene cluster; bioengineering; peptide synthetase module;  
 KM adenylation domain; hydroxyphenylglycine; HPG; antibiotic precursor;  
 KM chlorinate; lipdipeptide.  
 XX  
 OS Actinoplanes sp.  
 XX  
 PN WO200231155-A2.  
 XX  
 PD 18-APR-2002.  
 XX  
 PF 15-OCT-2001; 2001WO-CA01462.  
 XX  
 PR 13-OCT-2000; 2000US-239924P.  
 PR 12-APR-2001; 2001US-283296P.  
 PR 24-JUL-2001; 2001US-0910813.  
 XX  
 PA (ECOP-) ECOPRIA BIOSCIENCES INC.  
 PI Farnet CM, Zazopoulos E, Stafla A;  
 XX WPI; 2002-435445/46.  
 DR N-PSDB; AAL40781.  
 XX

Novel isolated ramoplanin biosynthetic pathway polypeptide useful for chemically modifying biological molecule that is a substrate for a polypeptide encoded by a ramoplanin biosynthesis gene cluster -

Claim 14; Page 136-137; 212pp; English.

The invention relates to an isolated ramoplanin biosynthetic pathway polypeptide selected from a polypeptide of open reading frames (ORF) 1-32. The isolated polypeptides are useful for chemically modifying a biological molecule that is a substrate for a polypeptide encoded by a ramoplanin biosynthesis gene cluster, by contacting the biological molecule with the isolated polypeptide, where the polypeptide chemically modifies the biological molecule. The method comprises contacting the biological molecule with at least two different polypeptides encoded by ramoplanin ORFs 1-31. The polypeptides are useful for directing the biosynthesis of the antibiotic ramoplanin in microorganisms. An isolated gene cluster comprising the ORFs is useful as a substrate for bioengineering of antibiotic structures. An isolated polypeptide or its encoding nucleic acid sequence is useful for generating derivatives of ramoplanin, for improving production or for producing variants of other antibiotics of the peptide class. The isolated polypeptides are useful for synthesis of ramoplanin in vivo or in vitro, as an adenylation domain in conjunction with other peptide synthetase modules and allowing the incorporation of Thr into a peptide antibiotic precursor, for modifying fatty acid structure and/or enhancing fatty acid incorporation into the peptide antibiotic structure, for production of an hydroxyphenylglycine (HPG)-containing peptide antibiotic, for enhancing secretion of ramoplanin or its variants and derivatives, for enhancing uptake of precursors for ramoplanin biosynthesis, for enhancing production of ramoplanin products or its variants or derivatives, to chlorinate HPG of a peptide antibiotic precursor, and for designing specific nucleoside probes and primers for identifying and isolating putative lipdipeptide-producing microorganisms. This sequence represents one of the ORF proteins of the ramoplanin producing Actinoplanes sp. microorganism of

CC the invention.  
 XX  
 SQ Sequence 333 AA;  
 QY 10 LALLTL 16  
 |||||  
 DB 64 LALLTL 70

Query Match  
 Best Local Similarity 100.0%; Pred. No. 95;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 45  
 ID ABB69595 standard; Protein; 361 AA.  
 XX ABB69595;  
 AC ABB69595;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 35577.  
 XX  
 KM Drosophila; developmental biology; cell signalling; insecticide;  
 KM pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI; 2001-656860/75.  
 DR N-PSDB; ABL13698.  
 XX

New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

Disclosure; SEQ ID NO 35577; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA sequences (ABU1840-ABU16175) and the encoded proteins (ABBS7737-ABBS72072).

The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 361 AA;  
 SQ

Query Match  
 Best Local Similarity 100.0%; Pred. No. 18+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 LALLTL 14  
 |||||  
 DB 231 LALLTL 237

RESULT 46  
ID AAU08679 standard; Protein, 381 AA.  
AAU08679 standard; Protein, 381 AA.  
AAU08679; 18-DEC-2001 (first entry)  
Human FCTR3a polypeptide sequence.  
Human, FCTRX; myelogenous leukaemia; carcinoma; melanoma; glioma; astrocytoma; congenital neonatal alloimmune thrombocytopenia; infection; neurological disorder; neurodegenerative disorders; nerve trauma; familial myelodysplastic syndrome; Charcot-Marie-Tooth neuropathy; demyelinating Gardner syndrome; familial myelodysplastic syndrome; mental health condition; immunological disorder; allergy; infertility; bronchial asthma; Avellino type eosinophilia; lung disease; deafness; reproductive disorder; reproductive disorder; glycoprotein Ia deficiency; demoid disease; turcot syndrome; liver cirrhosis; hepatitis C; viral; gastric disorders; pancreatic disease; Schistosoma mansoni infection; Spino cerebellar ataxia; Plasmodium falciparum parasitaemia; diabetes; Corneal dystrophy-Greenow type I; Corneal dystrophy-lattice type I; Reis-Bucklers corneal dystrophy; cytostatic; immunosuppressive; anti-allergic; antiaesthetic; antifertility; anti-inflammatory; antidiabetic; protozoicide; hepatotropic; virucide; ophthalmological; gynaecological; antinfertility; immunostimulant; auditory; haemostatic; gene therapy; FCTR3a; neurastin-like protein.  
Homo sapiens.  
MO20016747-A2.  
13-SEP-2001.  
05-MAR-2001; 2001MO-US07160.  
03-MAR-2000; 2000US-186592P.  
03-MAR-2000; 2000US-186718P.  
06-MAR-2000; 2000US-187293P.  
06-MAR-2000; 2000US-187294P.  
17-MAR-2000; 2000US-190400P.  
07-APR-2000; 2000US-196018P.  
03-JAN-2001; 2001US-259548P.  
(CURAGEN) CURAGEN CORP.  
Vernet CM, Fernandes E, Shimkera RA, Herrmann JL, Majumder K; MacDougall J, Mishra V, Mezes PS, Raetelli L; WPI; 2001-596837/67.  
N-PSDB; AAS14084.  
Novel polypeptides designated as FCTRX polypeptides, useful in detection, prevention and treatment of a broad range of pathological states -  
Claim 1; Page 33; 215pp; English.

CC Rejs-Bucklers corneal dystrophy. This sequence represents FCTR3a, a  
CC neurastin-like protein.  
SQ Sequence 381 AA;  
QY 10 LLLALLL 16  
DB 233 LLLALLL 239  
Query Match 5.1%; Score 7; DB 22; Length 381;  
Best Local Similarity 100.0%; Pred. No. 1,1e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0  
RESULT 47  
AAB61141  
ID AAB61141 standard; Protein; 381 AA.  
XX AAB61141;  
AC AAB61141;  
XX 30-MAR-2001 (first entry)  
XX Human NOV11 protein.  
DE Human NOV11 protein.  
XX Human; NOVX; antiinflammatory; cytostatic; neuroprotective;  
KW cerebroprotective; immunomodulator; vulnerary; vasotrophic; gene therapy;  
RV hyperplasia; tumour; xerostomatitis; psoriasis; Dupuytren's contracture;  
XX diabetes; rheumatoid arthritis; cerebral oedema; Alzheimer's disease.  
XX Homo sapiens.  
OS Homo sapiens.  
FN WO200075321-A2.  
PD 14-DEC-2000.  
XX 01-JUN-2000; 2000WO-US15303.  
PF 03-JUN-1999; 99US-0137322.  
PR 16-MAR-2000; 2000US-0189810.  
PR 22-MAR-2000; 2000US-0191158.  
PR 30-MAR-2000; 2000US-0193086.  
PR 31-MAY-2000; 2000US-0137322.  
XX (CURAGEN CORP.  
PA Shinkets RA, Fernandes B, Herman J, Vernet C;  
PI WPI; 2001-102403/11.  
DR N-PSDB; AAF27859.  
XX New NOVX polypeptides and polynucleotides, useful in gene therapy, as a  
PT diagnostic marker, protein therapeutic, antibody or small molecule drug  
PT target for treating immune, proliferative and metabolic diseases and  
PT wound healing -  
PS Claim 1; Page 42-44; 194pp; English.  
XX The present sequence is a new isolated polypeptide (NOVX). The NOVX  
CC polypeptides, NOVX nucleic acids, and anti-NOVX antibodies are useful for  
CC treating or preventing NOVX-associated disorders. They are also useful  
CC for determining the presence of or a predisposition to a disease  
CC associated with altered levels of the NOVX polypeptide or nucleic acid.  
CC These NOVX-associated disorders include hyperplasias, tumours,  
CC xerostomatitis, psoriasis, Dupuytren's contracture, diabetic complications,  
CC rheumatoid arthritis, cerebral lesions, diabetic neuropathies, cerebral  
CC oedema, senile dementia or Alzheimer's disease. The NOVX polynucleotides  
CC are especially useful in gene therapy. Specifically, NOVX is useful as  
CC a diagnostic marker or prognostic marker, protein therapeutic and  
CC antibody target or small molecule drug target to treat disorders in the  
CC immune response pathway, thyroid and metabolic diseases, bone metabolic  
CC disorders, diseases of the pancreas (e.g. diabetes or digestive  
CC disorders), proliferative diseases, or tissue regeneration and  
CC development (e.g. wound healing or treatment of burns).

XX SQ Sequence 381 AA;  
 Query Match 5.1%; Score 7; DB 22; Length 381;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 10 LLLAAILL 16  
 DB 233 LLLAAILL 239  
 RESULT 48  
 AAY37807  
 ID AAY37807 standard; Protein, 394 AA.  
 AC AAY37807;  
 DT 07-OCT-1999 (first entry)  
 DE Amino acid sequence of a Chlamydia trachomatis protein.  
 XX  
 KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
 KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;  
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;  
 KW Bartholinitis; pneumonia; venereal lymphogranulomatosis.  
 XX  
 OS Chlamydia trachomatis.  
 XX  
 PN MO9928475-A2.  
 PD 10-JUN-1999.  
 XX  
 PF 27-NOV-1998; 98WO-1B01939.  
 XX  
 PR 04-NOV-1998; 98US-0107077.  
 PR 28-NOV-1997; 97FR-0015041.  
 PR 17-DEC-1997; 97FR-0016034.  
 XX  
 PA (GIST ) GENSET.  
 XX  
 PI Griffiths R;  
 XX  
 DR WPI; 1999-371125/31.  
 XX  
 PT Genome sequence of Chlamydia trachomatis  
 PS Disclosure; Page 1383; 1755pp; English.  
 XX  
 CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome  
 CC of Chlamydia trachomatis (see AAY36754). The polypeptides can be used as  
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences  
 CC can also be used to control growth of the microorganism. Chlamydia  
 CC trachomatis is responsible for a large number of diseases, e.g. eye  
 CC diseases such as conventional trachoma, nonendemic trachoma,  
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as  
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,  
 CC Bartholinitis, Bartholinitis; pneumonia in breast feeding infants;  
 CC and venereal lymphogranulomatosis. The polypeptides of the invention  
 CC may be of use in treating these diseases.  
 XX  
 SQ Sequence 394 AA;  
 Query Match 5.1%; Score 7; DB 20; Length 394;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 QALQQLL 12  
 DB 57 QALQQLL 63  
 RESULT 49

AAB36700  
 ID AAB36700 standard; Protein, 415 AA.  
 XX  
 AC AAB36700;  
 DT 15-MAR-2001 (first entry)  
 DE Human tumour necrosis factor receptor LTBR protein SEQ ID NO:6.  
 XX  
 KW Human; tumour necrosis factor receptor 5; TRID; TNFR-5; TR5; nontropic;  
 KW TRAIL receptor without intracellular domain; diagnosis; cytostatic;  
 KW tumour necrosis factor related apoptosis inducing ligand; vasotropic;  
 KW immunosuppressive; neuroprotective; antiviral; antiinflammatory;  
 KW anticonvulsant; antiparasitic; cardiant; anti-HIV; antiparkinsonian;  
 KW gene therapy; restenosis; graft versus host disease; tumour; cancer;  
 KW apoptotic cell death related disease; autoimmune disorder;  
 KW cardiovascular disorder; viral infection.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200071150-A1.  
 PD 30-NOV-2000.  
 XX  
 PF 18-MAY-2000; 2000MO-US13515.  
 PR 20-MAY-1999; 99US-0135164.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Wei Y, Ruben SM, Gentz RL, Ni J;  
 XX  
 DR WPI; 2001-041051/05.  
 XX  
 PT Nucleic acid encoding a TRID polypeptide, also referred to as tumor  
 PT necrosis factor receptor 5, useful in the diagnosis, treatment or  
 PT prevention of cancer, autoimmune disorders and viral infection -  
 XX  
 PS Disclosure; Fig 2; 285pp; English.  
 XX  
 CC The present invention describes the human TRID protein (tumour necrosis  
 CC factor (TNF) related apoptosis inducing ligand (TRAIL) receptor without  
 CC intracellular domain, also referred to as tumour necrosis factor  
 CC receptor 5 (TNFR-5 or TR5)). TRID has cytostatic, immunosuppressive,  
 CC neurotropic, neuroprotective, antiviral, antiinflammatory, anticonvulsant,  
 CC antiparasitic, cardiant, anti-HIV, antiparkinsonian and vasotropic  
 CC activities, and can be used in gene therapy. The TRID polynucleotides  
 CC are useful for detecting complementary polynucleotides. TRID proteins and  
 CC polynucleotides are useful in the treatment of tumours, resistance to  
 CC parasite, bacteria and viruses, restenosis and graft versus host disease.  
 CC They are also useful for inducing proliferation of T-cells, endothelial  
 CC cells and certain haematopoietic cells, to regulate antiviral responses  
 CC and to prevent certain autoimmune diseases after stimulation of TRID by  
 CC an agonist or TRAIL binding facilitator. The antibodies which bind TRID  
 CC polypeptides are useful for treating and/or preventing diseases  
 CC associated with increased or decreased apoptotic cell death. The TRID  
 CC polynucleotides, proteins, antibodies, agonists and antagonists are  
 CC useful in the diagnosis, treatment or prevention of: (a) cancer;  
 CC (b) autoimmune disorders; (c) diseases associated with increased  
 CC apoptosis; (d) cardiovascular disorders; and (e) viral infection. The  
 CC present sequence represents a tumour necrosis factor receptor used in  
 CC comparison with TRID in the exemplification of the present invention.  
 XX  
 SQ Sequence 415 AA;  
 Query Match 5.1%; Score 7; DB 22; Length 415;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 10 LLLAAILL 16  
 DB 222 LLLAAILL 228

## RESULT 50

ABP96138 standard; Protein; 415 AA.

ID ABP96138

AC ABP96138;

XX 09-MAY-2003 (first entry)

XX Mouse lymphotoxin-beta receptor protein SEQ ID NO:20.

XX Human; tumour necrosis factor receptor 2 related protein variant;

XX TNFR2PV; cytosolic; immunosuppressive; antiasthmatic; gene therapy;

XX TNF signaling; cancer; inflammatory disorder; rheumatoid arthritis;

XX asthma; ulcerative colitis.

XX Mus musculus.

XX WO2003012037-A2.

XX 13-FEB-2003.

XX 24-JUL-2002; 2002WO-US23684.

XX 27-JUL-2001; 2001US-0917372.

XX (INCY-) INCYTE GENOMICS INC.

XX Lal PG, Warren BA;

XX MPI; 2003-247895/25.

XX New CDNA, useful for preparing a composition for treating a disease or

XX condition associated with increased TNF signaling e.g., cancer of the

XX prostate, ovary, gallbladder, breast, brain, liver or colon, or

XX rheumatoid arthritis, asthma

XX Disclosure; Fig 2A-C; 64pp; English.

XX The present invention describes human tumour necrosis factor receptor 2

XX related protein variant (TNFR2PV). TNFR2PV has cytostatic, antiasthmatic

XX and immunosuppressive activities, and can be used in gene therapy. The

XX TNFR2PV CDNA or protein sequences can be used for preparing a composition

XX for treating a disease or condition associated with increased TNF

XX signaling e.g., cancer of the prostate, ovary, gallbladder, breast,

XX brain, liver or colon, or inflammatory disorders, such as rheumatoid

XX arthritis, asthma or ulcerative colitis. The present sequence represents

XX a mouse lymphotoxin-beta receptor amino acid sequence, which is given

XX in comparison with human TNFR2PV in the exemplification of the present

XX invention.

XX Sequence 415 AA;

XX Query Match 5.1%; Score 7; DB 24; Length 415;

XX Best Local Similarity 100.0%; Pred.No. 1.2e+02;

XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLAAILL 16  
|||  
Db 222 LLAAILL 228Search completed: January 21, 2004, 12:26:34  
Job time : 43 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 12:27:27 ; Search time 33 Seconds  
(without alignments)  
848.907 Million cell updates/sec

Title: US-09-941-314-2

Perfect score: 137  
Sequence: 1 MAFPMQALQLLAITLTLMA.....VFAVPWFQYKIKKSCSSD 137

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 762491 seqs, 204481190 residues

Word size : 0

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09C\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIS

Result No.	Score	Query Match	Length	ID	Description
1	137	100.0	137	10	US-09-941-314-2 Sequence 2, Appl1
2	117	85.4	117	10	US-09-941-314-4 Sequence 4, Appl1
3	115	83.9	115	10	US-09-941-314-3 Sequence 3, Appl1
4	80	58.4	80	10	US-09-941-314-15 Sequence 15, Appl1
5	59	43.1	59	10	US-09-941-314-16 Sequence 16, Appl1
6	52	38.0	52	10	US-09-941-314-14 Sequence 14, Appl1
7	50	36.5	50	9	US-09-864-761-34822 Sequence 34822, A
8	50	36.5	50	9	US-09-864-761-48936 Sequence 48936, A
9	49	35.8	49	10	US-09-941-314-13 Sequence 13, Appl1
10	48	35.0	48	10	US-09-941-314-17 Sequence 17, Appl1
11	46	33.6	46	10	US-09-941-314-10 Sequence 10, Appl1
12	36	26.3	36	10	US-09-941-314-6 Sequence 6, Appl1
13	35	25.5	35	10	US-09-941-314-8 Sequence 8, Appl1
14	33	24.1	33	10	US-09-941-314-12 Sequence 12, Appl1
15	27	19.7	27	10	US-09-941-314-9 Sequence 9, Appl1

16	24	17.5	24	10	US-09-941-314-11	Sequence 11, Appl1
17	22	16.1	22	10	US-09-941-314-5	Sequence 5, Appl1
18	18	13.1	18	10	US-09-941-314-7	Sequence 7, Appl1
19	8	5.8	1851	12	US-10-369-493-6050	Sequence 6050, Ap
20	8	5.8	1966	12	US-10-369-493-6049	Sequence 6049, Ap
21	8	5.8	2848	12	US-10-369-493-6048	Sequence 6048, Ap
22	7	5.1	112	10	US-09-893-737-332	Sequence 322, App
23	7	5.1	184	12	US-10-312-273-375	Sequence 375, App
24	7	5.1	330	9	US-09-912-020-261	Sequence 261, App
25	7	5.1	332	12	US-10-369-493-5849	Sequence 5849, Ap
26	7	5.1	333	10	US-09-976-052-2	Sequence 2, Appl1
27	7	5.1	353	12	US-10-029-386-33896	Sequence 33896, A
28	7	5.1	381	11	US-09-808-602-6	Sequence 6, Appl1
29	7	5.1	381	11	US-09-977-032A-22	Sequence 22, Appl
30	7	5.1	381	11	US-09-800-198-6	Sequence 22, Appl
31	7	5.1	381	11	US-09-800-198-6	Sequence 22, Appl
32	7	5.1	381	12	US-09-977-751C-22	Sequence 22, Appl
33	7	5.1	381	12	US-09-977-639A-22	Sequence 22, Appl
34	7	5.1	381	12	US-09-977-819B-22	Sequence 22, Appl
35	7	5.1	415	9	US-09-826-212-6	Sequence 6, Appl1
36	7	5.1	415	9	US-09-907-372-20	Sequence 20, Appl
37	7	5.1	415	10	US-09-935-727-8	Sequence 8, Appl1
38	7	5.1	415	11	US-09-917-372-20	Sequence 20, Appl
39	7	5.1	415	15	US-10-186-643-6	Sequence 6, Appl1
40	7	5.1	417	15	US-10-156-761-10904	Sequence 10904, A
41	7	5.1	429	12	US-10-098-871-8	Sequence 8, Appl1
42	7	5.1	461	12	US-10-108-260A-3238	Sequence 3238, Ap
43	7	5.1	495	12	US-10-094-749-2349	Sequence 2349, Ap
44	7	5.1	506	11	US-09-769-787-80	Sequence 80, Appl
45	7	5.1	661	14	US-10-114-893-10	Sequence 10, Appl
46	7	5.1	917	9	US-09-861-801-2	Sequence 2, Appl1
47	7	5.1	917	12	US-10-170-789-28	Sequence 28, Appl
48	7	5.1	917	12	US-10-108-260A-2966	Sequence 2966, Ap
49	7	5.1	1844	12	US-10-262-056-53	Sequence 53, Appl
50	7	5.1	1844	15	US-10-242-056-53	Sequence 53, Appl
51	7	5.1	2504	9	US-09-817-514A-8	Sequence 8, Appl1
52	7	5.1	2504	12	US-10-262-794A-12	Sequence 12, Appl
53	7	5.1	2504	15	US-10-242-056-12	Sequence 12, Appl
54	7	5.1	2551	12	US-10-144-194A-80	Sequence 80, Appl
55	7	5.1	2631	12	US-10-144-194A-82	Sequence 82, Appl
56	7	5.1	2744	10	US-09-808-602-13	Sequence 13, Appl
57	7	5.1	2744	11	US-09-800-198-13	Sequence 8, Appl1
58	7	5.1	2733	11	US-09-808-602-8	Sequence 8, Appl1
59	7	5.1	2733	10	US-09-800-198-8	Sequence 80, Appl
60	7	5.1	2764	10	US-09-808-602-80	Sequence 80, Appl
61	7	5.1	2764	11	US-09-800-198-68	Sequence 68, Appl
62	7	5.1	2765	10	US-09-808-602-84	Sequence 84, Appl
63	7	5.1	2765	11	US-09-800-198-72	Sequence 72, Appl
64	7	5.1	2765	15	US-09-800-198-72	Sequence 12, Appl
65	6	4.4	14	15	US-10-283-838-12	Sequence 12, Appl
66	6	4.4	30	9	US-09-760-008A-7	Sequence 7, Appl1
67	6	4.4	30	12	US-10-318-966-7	Sequence 7, Appl1
68	6	4.4	30	15	US-10-192-294-7	Sequence 7, Appl1
69	6	4.4	31	15	US-10-001-883-131	Sequence 131, App
70	6	4.4	34	12	US-10-231-417-543	Sequence 543, App
71	6	4.4	46	12	US-10-082-828A-259	Sequence 259, App
72	6	4.4	51	9	US-09-864-761-47818	Sequence 47818, A
73	6	4.4	54	15	US-10-137-036-130	Sequence 130, App
74	6	4.4	56	15	US-10-083-357-834	Sequence 834, App
75	6	4.4	57	15	US-10-097-065-173	Sequence 173, App
	6	4.4	62	9	US-09-864-761-48278	Sequence 48278, A

#### ALIGNMENTS

RESULT 1  
US-09-941-314-2  
; Application US/09941314  
; Patent No. US20020142396A1  
; GENERAL INFORMATION:  
; APPLICANT: ZymoGenetics, Inc.  
; TITLE OF INVENTION: Mammalian Cystatin-B and Its Use to  
; Inhibit Cancer Procoagulant Protein

FILE REFERENCE: 00-81PC  
CURRENT APPLICATION NUMBER: US/09/941,314  
CURRENT FILING DATE: 2001-08-29  
PRIOR APPLICATION NUMBER: 60/230,230  
PRIOR FILING DATE: 2001-09-01  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 137  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-941-314-2

Query Match 100.0%; Score 137; DB 10; Length 137;  
Best Local Similarity 100.0%; Pred. No. 2,4e-126;  
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAEPWQALQMLLALITLMLPYQARKTFLSVHEVAVENYAKDSLQWITDQYNKSSDD 60  
Db 1 MAEPWQALQMLLALITLMLPYQARKTFLSVHEVAVENYAKDSLQWITDQYNKSSDD 60  
Qy 61 KYHFRIRFVLKVQROVTDHLEHYHNVEMQWTTCKPRTTNCVPORELIHKQVNCFFSVFA 120  
Db 61 KYHFRIRFVLKVQROVTDHLEHYHNVEMQWTTCKPRTTNCVPORELIHKQVNCFFSVFA 120

Qy 121 VPMFEQYKILNKSSSD 137  
Db 121 VPMFEQYKILNKSSSD 137

## RESULT 2

US-09-941-314-4  
Sequence 4, Application US/09941314  
Patent No. US20020142396A1  
GENERAL INFORMATION:  
APPLICANT: ZymoGenetics, Inc.  
TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to  
TITLE OF INVENTION: Inhibit Cancer Procoagulant Protein  
FILE REFERENCE: 00-81PC  
CURRENT APPLICATION NUMBER: US/09/941,314  
CURRENT FILING DATE: 2001-08-29  
PRIOR APPLICATION NUMBER: 60/230,230  
PRIOR FILING DATE: 2001-09-01  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 4  
LENGTH: 117  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-941-314-4

Query Match 85.4%; Score 117; DB 10; Length 117;  
Best Local Similarity 100.0%; Pred. No. 8.1e-107;  
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 21 LPYQARKTFLSVHEVAVENYAKDSLQWITDQYNKSSDDKYHFRIRFVLKVQROVTDHL 80  
Db 1 LPYQARKTFLSVHEVAVENYAKDSLQWITDQYNKSSDDKYHFRIRFVLKVQROVTDHL 60  
Qy 81 EYHNVEMQWTTCKPRTTNCVPORELIHKQVNCFFSVFAVPMFEQYKILNKSSSD 137  
Db 61 EYHNVEMQWTTCKPRTTNCVPORELIHKQVNCFFSVFAVPMFEQYKILNKSSSD 117

## RESULT 3

US-09-941-314-3  
Sequence 3, Application US/09941314  
Patent No. US20020142396A1  
GENERAL INFORMATION:  
APPLICANT: ZymoGenetics, Inc.  
TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to  
TITLE OF INVENTION: Inhibit Cancer Procoagulant Protein  
FILE REFERENCE: 00-81PC

CURRENT APPLICATION NUMBER: US/09/941,314  
CURRENT FILING DATE: 2001-08-29  
PRIOR APPLICATION NUMBER: 60/230,230  
PRIOR FILING DATE: 2001-09-01  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 115  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-941-314-3

Query Match 83.9%; Score 115; DB 10; Length 115;  
Best Local Similarity 100.0%; Pred. No. 7.3e-105;  
Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 YQARKTFLSVHEVAVENYAKDSLQWITDQYNKSDKTHFRIFVLKVQROVTDHLEY 82  
Db 1 YQARKTFLSVHEVAVENYAKDSLQWITDQYNKSDKTHFRIFVLKVQROVTDHLEY 60  
Qy 83 HNVEMQWTTCKPRTTNCVPORELIHKQVNCFFSVFAVPMFEQYKILNKSSSD 137  
Db 61 HNVEMQWTTCKPRTTNCVPORELIHKQVNCFFSVFAVPMFEQYKILNKSSSD 115

## RESULT 4

US-09-941-314-15  
Sequence 15, Application US/09941314  
Patent No. US20020142396A1  
GENERAL INFORMATION:  
APPLICANT: ZymoGenetics, Inc.  
TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to  
TITLE OF INVENTION: Inhibit Cancer Procoagulant Protein  
FILE REFERENCE: 00-81PC  
CURRENT APPLICATION NUMBER: US/09/941,314  
CURRENT FILING DATE: 2001-08-29  
PRIOR APPLICATION NUMBER: 60/230,230  
PRIOR FILING DATE: 2001-09-01  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15  
LENGTH: 80  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-941-314-15

Query Match 58.4%; Score 80; DB 10; Length 80;  
Best Local Similarity 100.0%; Pred. No. 1e-70;  
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 53 QYNKESDCKHFRIFVLKVQROVTDHLEHYHNVEMQWTTCKPRTTNCVPORELIHKQV 112  
Db 1 QYNKESDCKHFRIFVLKVQROVTDHLEHYHNVEMQWTTCKPRTTNCVPORELIHKQV 60  
Qy 113 NCFPSVFAVPMFEQYKILNK 132  
Db 61 NCFPSVFAVPMFEQYKILNK 80

## RESULT 5

US-09-941-314-16  
Sequence 16, Application US/09941314  
Patent No. US20020142396A1  
GENERAL INFORMATION:  
APPLICANT: ZymoGenetics, Inc.  
TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to  
TITLE OF INVENTION: Inhibit Cancer Procoagulant Protein  
FILE REFERENCE: 00-81PC  
CURRENT APPLICATION NUMBER: US/09/941,314  
CURRENT FILING DATE: 2001-08-29  
PRIOR APPLICATION NUMBER: 60/230,230  
PRIOR FILING DATE: 2001-09-01  
NUMBER OF SEQ ID NOS: 19

```
SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-941-314-16

Query Match      43.1%; Score 59; DB 10; Length 59;
Best Local Similarity 100.0%; Pred. No. 2.8e-50;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 RQVTLHLVHLNVEMQTTCCQKETTNCVQERLHKQVNCFFSVAVPMFEQYKLNK 132
DB 1 RQVTLHLVHLNVEMQTTCCQKETTNCVQERLHKQVNCFFSVAVPMFEQYKLNK 59

RESULT 6
US-09-941-314-14
; Sequence 14, Application US/09941314
; Patent No. US20020142396A1
; GENERAL INFORMATION:
; APPLICANT: ZymoGenetics, Inc.
; TITLE OF INVENTION: Mammalian Cysteatin-8 and its use to
; FILE REFERENCE: 00-81PC
; CURRENT APPLICATION NUMBER: US/09/941,314
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/230,230
; PRIOR FILING DATE: 2001-09-01
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-941-314-14

Query Match      38.0%; Score 52; DB 10; Length 52;
Best Local Similarity 100.0%; Pred. No. 1.8e-43;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 KESDDKXHFRIFRVLKQROVTDHLEHLNVEMQTTCCQKETTNCVQERE 107
DB 1 KESDDKXHFRIFRVLKQROVTDHLEHLNVEMQTTCCQKETTNCVQERE 52

RESULT 7
US-09-864-761-34822
; Sequence 34822, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
```

```
PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34822
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL109954.10
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.2
; OTHER INFORMATION: EST HUMAN HIT: A1200857.1, EVALU 5.00e-23
; OTHER INFORMATION: SWISSPROT HIT: O60676, EVALU 1.00e-01
US-09-864-761-34822

Query Match      36.5%; Score 50; DB 9; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.6e-41;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 RKTFLSVHEVMAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKQRO 75
DB 1 RKTFLSVHEVMAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKQRO 50

RESULT 8
US-09-864-761-48936
; Sequence 48936, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
```

```
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263.6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 48936
/ LENGTH: 50
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AL096677.18
/ OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.96
/ OTHER INFORMATION: EST_HUMAN HIT: A1200857.1, EVALU6 5.00e-23
/ OTHER INFORMATION: SWISSPROT HIT: O60676, EVALU6 1.00e-01
/ US-09-864-761-48936
```

Query Match 36.5%; Score 50; DB 9; Length 50;

Best Local Similarity 100.0%; Pred. No. 1.6e-41; Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 26 RKTFLSVHEVMAVENVAKDSLQWITDQYNKESDDKXHFRIFRVLKQVQ 75
Db 1 RKTFLSVHEVMAVENVAKDSLQWITDQYNKESDDKXHFRIFRVLKQVQ 50
```

```
RESULT 9
/ US-09-941-314-13
/ Sequence 13, Application US/09941314
/ Patent No. US20020142396A1
/ GENERAL INFORMATION:
/ APPLICANT: ZymoGenetics, Inc.
/ TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to
/ TITLE OF INVENTION: Inhibit Cancer Procoagulant Protein
/ FILE REFERENCE: 00-81PC
/ CURRENT APPLICATION NUMBER: US/09/941,314
/ PRIOR FILING DATE: 2001-08-29
/ PRIOR APPLICATION NUMBER: 60/230,230
/ NUMBER OF SEQ ID NOS: 19
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 13
/ LENGTH: 49
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ US-09-941-314-13
```

Query Match 35.8%; Score 49; DB 10; Length 49;

Best Local Similarity 100.0%; Pred. No. 1.5e-40; Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 52 DQYNKESDDKXHFRIFRVLKQVQVTDHLEVHLNVEMQWTTCKPFTTN 100
Db 1 DQYNKESDDKXHFRIFRVLKQVQVTDHLEVHLNVEMQWTTCKPFTTN 49
```

```
RESULT 10
/ US-09-941-314-17
/ Sequence 17, Application US/09941314
/ Patent No. US20020142396A1
/ GENERAL INFORMATION:
/ APPLICANT: ZymoGenetics, Inc.
/ TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to
/ TITLE OF INVENTION: Inhibit Cancer Procoagulant Protein
/ FILE REFERENCE: 00-81PC
/ CURRENT APPLICATION NUMBER: US/09/941,314
/ PRIOR FILING DATE: 2001-08-29
/ PRIOR APPLICATION NUMBER: 60/230,230
/ NUMBER OF SEQ ID NOS: 19
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 17
/ LENGTH: 48
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ US-09-941-314-17
```

Query Match 35.0%; Score 48; DB 10; Length 48;

Best Local Similarity 100.0%; Pred. No. 1.4e-39; Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 85 NVEMQWTTCKPFTTNVCPQEREHLKQVNCFPVFAVFPPEQYKILNK 132
Db 1 NVEMQWTTCKPFTTNVCPQEREHLKQVNCFPVFAVFPPEQYKILNK 48
```

```
RESULT 11
/ US-09-941-314-10
/ Sequence 10, Application US/09941314
/ Patent No. US20020142396A1
/ GENERAL INFORMATION:
/ APPLICANT: ZymoGenetics, Inc.
/ TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to
/ TITLE OF INVENTION: Inhibit Cancer Procoagulant Protein
/ FILE REFERENCE: 00-81PC
/ CURRENT APPLICATION NUMBER: US/09/941,314
/ PRIOR FILING DATE: 2001-08-29
/ PRIOR APPLICATION NUMBER: 60/230,230
/ NUMBER OF SEQ ID NOS: 19
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 10
/ LENGTH: 46
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ US-09-941-314-10
```

Query Match 33.6%; Score 46; DB 10; Length 46;

Best Local Similarity 100.0%; Pred. No. 1.2e-37; Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 55 NKESDDKXHFRIFRVLKQVQVTDHLEVHLNVEMQWTTCKPFTTN 100
Db 1 NKESDDKXHFRIFRVLKQVQVTDHLEVHLNVEMQWTTCKPFTTN 46
```

```
RESULT 12
/ US-09-941-314-6
/ Sequence 6, Application US/09941314
```



RESULT 17  
 US-09-941-314-5  
 ; Sequence 5, Application US/09941314  
 ; Patent No. US20020142396A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zymogenetics, Inc.  
 ; TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to  
 ; FILE REFERENCE: 00-81PC  
 ; CURRENT APPLICATION NUMBER: US/09/941,314  
 ; CURRENT FILING DATE: 2001-08-29  
 ; PRIOR APPLICATION NUMBER: 60/230,230  
 ; PRIOR FILING DATE: 2001-09-01  
 ; NUMBER OF SEQ ID NOS: 19  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 5  
 ; LENGTH: 22  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-941-314-5

Query Match 16.1%; Score 22; DB 10; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-14;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 QARKKFLSVHEVMAVENYARD 45  
 DB 1 QARKKFLSVHEVMAVENYARD 22

RESULT 18  
 US-09-941-314-7  
 ; Sequence 7, Application US/09941314  
 ; Patent No. US20020142396A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zymogenetics, Inc.  
 ; TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to  
 ; FILE REFERENCE: 00-81PC  
 ; CURRENT APPLICATION NUMBER: US/09/941,314  
 ; CURRENT FILING DATE: 2001-08-29  
 ; PRIOR APPLICATION NUMBER: 60/230,230  
 ; PRIOR FILING DATE: 2001-09-01  
 ; NUMBER OF SEQ ID NOS: 19  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 7  
 ; LENGTH: 18  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-941-314-7

Query Match 13.1%; Score 18; DB 10; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-10;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 KDSIQWITTDQYNKESDDK 61  
 DB 1 KDSIQWITTDQYNKESDDK 18

RESULT 19  
 US-10-369-493-6050  
 ; Sequence 6050, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 6050  
 ; LENGTH: 1851  
 ; TYPE: PRT  
 ; ORGANISM: Caenorhabditis elegans  
 US-10-369-493-6050

Query Match 5.8%; Score 8; DB 12; Length 1851;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLLIAIILL 16  
 DB 1093 QLLIAIILL 1100

RESULT 20  
 US-10-369-493-6049  
 ; Sequence 6049, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 6049  
 ; LENGTH: 1966  
 ; TYPE: PRT  
 ; ORGANISM: Caenorhabditis elegans  
 US-10-369-493-6049

Query Match 5.8%; Score 8; DB 12; Length 1966;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLLIAIILL 16  
 DB 1093 QLLIAIILL 1100

RESULT 21  
 US-10-369-493-6048  
 ; Sequence 6048, Application US/10369493  
 ; Publication No. US20030233675A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cao, Yongwei  
 ; APPLICANT: Hinkle, Gregory J.  
 ; APPLICANT: Slater, Steven C.  
 ; APPLICANT: Goldman, Barry S.  
 ; APPLICANT: Chen, Xianfeng  
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 ; FILE REFERENCE: 38-10(52052)B  
 ; CURRENT APPLICATION NUMBER: US/10/369,493  
 ; CURRENT FILING DATE: 2003-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/360,039  
 ; PRIOR FILING DATE: 2002-02-21  
 ; NUMBER OF SEQ ID NOS: 47374  
 ; SEQ ID NO 6048

LENGTH: 2848  
 TYPE: PRT  
 ORGANISM: Caenorhabditis elegans  
 US-10-369-493-6048

Query Match  
 Best Local Similarity 100.0%; Pred. No. 77;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 QLLLAAILL 16  
 Db 1093 QLLLAAILL 1100

RESULT 22  
 US-09-893-737-322  
 Sequence 322, Application US/09893737  
 Patent No. US20020110855A1  
 GENERAL INFORMATION:

APPLICANT: Sheppard, Paul O.  
 APPLICANT: Presnell, Scott R.  
 TITLE OF INVENTION: MAMMALIAN SECRETED PROTEINS  
 FILE REFERENCE: 00-41  
 CURRENT APPLICATION NUMBER: US/09/893,737  
 CURRENT FILING DATE: 2001-06-28  
 PRIOR APPLICATION NUMBER: US 60/215,446  
 PRIOR FILING DATE: 2000-06-30  
 NUMBER OF SEQ ID NOS: 329  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 322  
 LENGTH: 112  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-893-737-322

Query Match  
 Best Local Similarity 100.0%; Pred. No. 43;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ALQLLLA 13  
 Db 12 ALQLLLA 18

RESULT 23  
 US-10-312-273-375  
 Sequence 375, Application US/10312273  
 Publication No. US20040005667A1  
 GENERAL INFORMATION:

APPLICANT: CHIRON SPA  
 TITLE OF INVENTION: IMMUNISATION AGAINST CHLAMYDIA PNEUMONIAE  
 FILE REFERENCE: P025035W0  
 CURRENT APPLICATION NUMBER: US/10/312,273  
 CURRENT FILING DATE: 2002-12-20  
 PRIOR APPLICATION NUMBER: 0016363.4  
 PRIOR FILING DATE: 2000-07-03  
 PRIOR APPLICATION NUMBER: 0017047.2  
 PRIOR FILING DATE: 2000-07-11  
 PRIOR APPLICATION NUMBER: 0017983.8  
 PRIOR FILING DATE: 2000-07-21  
 PRIOR APPLICATION NUMBER: 0019368.0  
 PRIOR FILING DATE: 2000-08-07  
 PRIOR APPLICATION NUMBER: 0020440.4  
 PRIOR FILING DATE: 2000-08-18  
 PRIOR APPLICATION NUMBER: 0022583.9  
 PRIOR FILING DATE: 2000-09-14  
 PRIOR APPLICATION NUMBER: 0027549.5  
 PRIOR FILING DATE: 2000-11-10  
 PRIOR APPLICATION NUMBER: 0031706.5  
 PRIOR FILING DATE: 2000-12-22  
 NUMBER OF SEQ ID NOS: 664  
 SOFTWARE: SeqWin99, version 1.02  
 SEQ ID NO 375

LENGTH: 184  
 TYPE: PRT  
 ORGANISM: Chlamydia pneumoniae  
 US-10-312-273-375

Query Match  
 Best Local Similarity 100.0%; Pred. No. 66;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 RKKTFPLS 32  
 Db 26 RKKTFPLS 32

RESULT 24  
 US-09-912-020-261  
 Sequence 261, Application US/09912020  
 Patent No. US20020045592A1  
 GENERAL INFORMATION:

APPLICANT: Zyskind, Judith  
 APPLICANT: Ohlsen, Karl L.  
 APPLICANT: Trawick, John  
 APPLICANT: Forsyth, R. Allyn  
 APPLICANT: Froelich, Jamie M.  
 APPLICANT: Carr, Grant J.  
 APPLICANT: Yamamoto, Robert T.  
 APPLICANT: Xu, H. Howard  
 TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN  
 FILE REFERENCE: ELITRA.001DV1  
 CURRENT APPLICATION NUMBER: US/09/912,020  
 CURRENT FILING DATE: 2001-07-23  
 PRIOR APPLICATION NUMBER: 09/492,709  
 PRIOR FILING DATE: 2000-01-27  
 PRIOR APPLICATION NUMBER: 60/117,405  
 PRIOR FILING DATE: 1999-01-27  
 NUMBER OF SEQ ID NOS: 485  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 261  
 LENGTH: 330  
 TYPE: PRT  
 ORGANISM: B. Coll  
 US-09-912-020-261

Query Match  
 Best Local Similarity 100.0%; Pred. No. 116+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LAILLTL 18  
 Db 88 LAILLTL 94

RESULT 25  
 US-10-369-493-5849  
 Sequence 5849, Application US/10369493  
 Publication No. US2003023675A1  
 GENERAL INFORMATION:

APPLICANT: Cao, Yongwei  
 APPLICANT: Hinkle, Gregory J.  
 APPLICANT: Slater, Steven C.  
 APPLICANT: Goldman, Barry S.  
 APPLICANT: Chen, Xianfeng  
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 FILE REFERENCE: 38-10(52052)B  
 CURRENT APPLICATION NUMBER: US/10/369,493  
 CURRENT FILING DATE: 2003-02-28  
 PRIOR APPLICATION NUMBER: US 60/360,039  
 PRIOR FILING DATE: 2002-02-21  
 NUMBER OF SEQ ID NOS: 47374  
 SEQ ID NO 5849  
 LENGTH: 332

TYPE: PRT  
ORGANISM: Caenorhabditis elegans  
US-10-369-493-5849

Query Match  
Best Local Similarity 100.0%; Score 7; DB 12; Length 332;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 RIRFVL 70  
Db 232 RIRFVL 238

RESULT 26  
US-09-976-059-2  
Sequence 2, Application US/09976059  
Patent No. US20020164747A1  
GENERAL INFORMATION:  
APPLICANT: Farnet, Chris  
APPLICANT: Zazopoulos, Emmanuel  
TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin  
FILE REFERENCE: 3019-PCT  
CURRENT APPLICATION NUMBER: US/09/976,059  
CURRENT FILING DATE: 2001-10-15  
NUMBER OF SEQ ID NOS: 34  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 2  
LENGTH: 333  
TYPE: PRT  
ORGANISM: Actinoplanes sp.  
US-09-976-059-2

Query Match  
Best Local Similarity 100.0%; Score 7; DB 10; Length 333;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 LLLAILL 16  
Db 64 LLLAILL 70

RESULT 27  
US-10-029-386-33896  
Sequence 33896, Application US/10029386  
Publication No. US20030194704A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hantzel, David K.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C  
TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
FILE REFERENCE: AEOMICA-X-2  
CURRENT APPLICATION NUMBER: US/10/029,386  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 34288  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 33896  
LENGTH: 353  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO Z93016.1  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3  
OTHER INFORMATION: SWISSPROT HIT: P15387, EVALU8 5.00e-59  
US-10-029-386-33896

Query Match  
Best Local Similarity 100.0%; Score 7; DB 12; Length 353;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 65 RIRFVLK 71  
Db 65 RIRFVLK 71

Db 331 RIRFVLK 337

RESULT 28  
US-09-808-602-6  
Sequence 6, Application US/09808602  
Patent No. US20020155115A1  
GENERAL INFORMATION:  
APPLICANT: Vermet, Corine A  
APPLICANT: Fernandes, Elma  
APPLICANT: Shimkets, Richard A  
APPLICANT: Herrman, John L  
APPLICANT: Majumder, Kumud  
APPLICANT: Mishra, Vishnu  
APPLICANT: Mezei, Peter S  
APPLICANT: MacDougall, John  
TITLE OF INVENTION: No. US20020155115A1 Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 15966-697 CIP  
CURRENT APPLICATION NUMBER: US/09/808,602  
CURRENT FILING DATE: 2001-03-14  
PRIOR APPLICATION NUMBER: 09/800,198  
PRIOR FILING DATE: 2001-03-05  
PRIOR APPLICATION NUMBER: 60/186,596  
PRIOR FILING DATE: 2000-03-03  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 6  
LENGTH: 381  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-808-602-6

Query Match  
Best Local Similarity 100.0%; Score 7; DB 10; Length 381;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 LLLAILL 16  
Db 233 LLLAILL 239

RESULT 29  
US-09-977-418-22  
Sequence 22, Application US/09977418  
Publication No. US20030027158A1  
GENERAL INFORMATION:  
APPLICANT: Shimkets et al  
TITLE OF INVENTION: No. US20030027158A1 polynucleotides and polypeptides encoded the  
FILE REFERENCE: 15966-552  
CURRENT APPLICATION NUMBER: US/09/977,418  
CURRENT FILING DATE: 2001-10-15  
PRIOR APPLICATION NUMBER: 09/584,411  
PRIOR FILING DATE: 2000-05-31  
PRIOR APPLICATION NUMBER: USSN 60/189,810  
PRIOR FILING DATE: 2000-03-16  
PRIOR APPLICATION NUMBER: USSN 60/191,158  
PRIOR FILING DATE: 2000-03-22  
PRIOR APPLICATION NUMBER: USSN 60/193,086  
PRIOR FILING DATE: 2000-03-30  
PRIOR APPLICATION NUMBER: USSN 60/201,388  
PRIOR FILING DATE: 2000-05-03  
NUMBER OF SEQ ID NOS: 93  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 22  
LENGTH: 381  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-977-418-22

Query Match  
Best Local Similarity 100.0%; Score 7; DB 11; Length 381;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLALL 16  
Db 233 LLLALL 239

## RESULT 30

US-09-977-033A-22  
; Sequence 22, Application US/09977033A  
; Publication No. US2003082554A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Herman, John  
; APPLICANT: Vernet, Corine  
; TITLE OF INVENTION: No. US2003082554A1el nucleic acid sequences encoding human KIAA  
; TITLE OF INVENTION: protein-like and human protein PRO28-11ke  
; FILE REFERENCE: 15966-552 CON-S24  
; CURRENT APPLICATION NUMBER: US/09/977,033A  
; PRIOR FILING DATE: 2001-10-15  
; PRIOR APPLICATION NUMBER: 60/137,322  
; PRIOR FILING DATE: 1999-06-03  
; PRIOR APPLICATION NUMBER: 60/189,810  
; PRIOR FILING DATE: 2000-03-16  
; PRIOR APPLICATION NUMBER: 60/191,158  
; PRIOR FILING DATE: 2000-03-22  
; PRIOR APPLICATION NUMBER: 60/193,086  
; PRIOR FILING DATE: 2000-03-30  
; PRIOR APPLICATION NUMBER: 60/201,388  
; PRIOR FILING DATE: 2000-05-03  
; PRIOR APPLICATION NUMBER: 09/584,411  
; PRIOR FILING DATE: 2000-05-31  
; NUMBER OF SEQ ID NOS: 92  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 22  
; LENGTH: 381  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-977-033A-22

Query Match 5.1%; Score 7; DB 11; Length 381;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLALL 16  
Db 233 LLLALL 239

## RESULT 31

US-09-800-198-6  
; Sequence 6, Application US/09800198  
; Publication No. US2003087816A1  
; GENERAL INFORMATION:  
; APPLICANT: Vernet, Cornie AM  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Hermann, John L  
; APPLICANT: Majumder, Kumud  
; APPLICANT: Mishra, Vishna  
; APPLICANT: Mezei, Peter S  
; APPLICANT: Rastelli, Luca  
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME  
; FILE REFERENCE: 15966-697  
; CURRENT APPLICATION NUMBER: US/09/800,198  
; PRIOR FILING DATE: 2001-03-05  
; PRIOR APPLICATION NUMBER: 60/186,596  
; PRIOR FILING DATE: 2000-03-03  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 381  
; TYPE: PRT

; ORGANISM: Homo sapiens  
US-09-800-198-6

Query Match 5.1%; Score 7; DB 11; Length 381;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLALL 16  
Db 233 LLLALL 239

## RESULT 32

US-09-977-751C-22  
; Sequence 22, Application US/09977751C  
; Publication No. US20030134430A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Herman, John  
; APPLICANT: Vernet, Corine  
; TITLE OF INVENTION: No. US20030134430A1el Amino Acid Sequences for Human Caenorhabditis  
; TITLE OF INVENTION: Polypeptides.  
; FILE REFERENCE: 15966-552 CON S-40  
; CURRENT APPLICATION NUMBER: US/09/977,751C  
; PRIOR FILING DATE: 2001-10-15  
; PRIOR APPLICATION NUMBER: 60/137,322  
; PRIOR FILING DATE: 1999-06-03  
; PRIOR APPLICATION NUMBER: 60/189,810  
; PRIOR FILING DATE: 2000-03-16  
; PRIOR APPLICATION NUMBER: 60/191,158  
; PRIOR FILING DATE: 2000-03-22  
; PRIOR APPLICATION NUMBER: 60/193,086  
; PRIOR FILING DATE: 2000-03-30  
; PRIOR APPLICATION NUMBER: 60/201,388  
; PRIOR FILING DATE: 2000-05-03  
; PRIOR APPLICATION NUMBER: 09/584,411  
; PRIOR FILING DATE: 2000-05-31  
; NUMBER OF SEQ ID NOS: 92  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 22  
; LENGTH: 381  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-977-751C-22

Query Match 5.1%; Score 7; DB 12; Length 381;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLALL 16  
Db 233 LLLALL 239

## RESULT 33

US-09-977-639A-22  
; Sequence 22, Application US/09977639A  
; Publication No. US20030199103A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Fernandes, Elma  
; APPLICANT: Herman, John  
; APPLICANT: Vernet, Corine  
; TITLE OF INVENTION: No. US20030199103A1el amino acid sequences for human epidermal grc  
; TITLE OF INVENTION: Polypeptides.  
; FILE REFERENCE: 15966-552 CON-S34  
; CURRENT APPLICATION NUMBER: US/09/977,639A  
; PRIOR FILING DATE: 2001-10-15  
; PRIOR APPLICATION NUMBER: 60/137,322  
; PRIOR FILING DATE: 1999-06-03  
; PRIOR APPLICATION NUMBER: 60/189,810  
; PRIOR FILING DATE: 2000-03-16

PRIOR APPLICATION NUMBER: 60/191,158  
PRIOR FILING DATE: 2000-03-22  
PRIOR APPLICATION NUMBER: 60193,086  
PRIOR FILING DATE: 2000-03-30  
PRIOR APPLICATION NUMBER: 60/201,388  
PRIOR FILING DATE: 2000-05-03  
PRIOR APPLICATION NUMBER: 09/584,411  
PRIOR FILING DATE: 2000-05-31  
NUMBER OF SEQ ID NOS: 92  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 22  
LENGTH: 381  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-977-639A-22

Query Match 5.1%; Score 7; DB 12; Length 381;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 LLLAIIIL 16  
Db 233 LLLAIIIL 239

RESULT 34  
US-09-977-819B-22

Sequence 22, Application US/09977819B  
Publication No. US20040002134A1  
GENERAL INFORMATION:  
APPLICANT: Shinketsu, Richard A  
APPLICANT: Fernandes, Elma  
APPLICANT: Herrman, John  
TITLE OF INVENTION: No. US20040002134A1 nucleic acid sequences encoding human KIAAC  
TITLE OF INVENTION: protein-like and human protein PRO228-1like  
FILE REFERENCE: 15966-552 CON-826  
CURRENT APPLICATION NUMBER: US/09/977,819B  
CURRENT FILING DATE: 2001-10-15  
PRIOR APPLICATION NUMBER: 60/137,322  
PRIOR FILING DATE: 1999-06-03  
PRIOR APPLICATION NUMBER: 60/189,810  
PRIOR FILING DATE: 2000-03-16  
PRIOR APPLICATION NUMBER: 60/191,158  
PRIOR FILING DATE: 2000-03-22  
PRIOR APPLICATION NUMBER: 60193,086  
PRIOR FILING DATE: 2000-03-30  
PRIOR APPLICATION NUMBER: 60/201,388  
PRIOR FILING DATE: 2000-05-03  
PRIOR APPLICATION NUMBER: 09/584,411  
PRIOR FILING DATE: 2000-05-31  
NUMBER OF SEQ ID NOS: 92  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 22  
LENGTH: 381  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-977-819B-22

Query Match 5.1%; Score 7; DB 12; Length 381;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 LLLAIIIL 16  
Db 233 LLLAIIIL 239

RESULT 35  
US-09-826-212-6  
Sequence 6, Application US/09826212  
Patent No. US20010021516A1

GENERAL INFORMATION:  
APPLICANT: Wei, Yang-Fei  
APPLICANT: Gentz, Reiner  
APPLICANT: Ruden, Steven  
APPLICANT: Ni, Jian  
TITLE OF INVENTION: Tumor Necrosis Factor Receptor 5  
FILE REFERENCE: 1488,1280006  
CURRENT APPLICATION NUMBER: US/09/826,212  
CURRENT FILING DATE: 2001-04-05  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 6  
LENGTH: 415  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-826-212-6

Query Match 5.1%; Score 7; DB 9; Length 415;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 LLLAIIIL 16  
Db 222 LLLAIIIL 228

RESULT 36  
US-09-907-372-20

Sequence 20, Application US/09907372  
Patent No. US20020068242A1  
GENERAL INFORMATION:  
APPLICANT: Lal, Preeti G.  
APPLICANT: Marten, Bridget A.  
TITLE OF INVENTION: TNF RECEPTOR 2 RELATED PROTEIN VARIANT  
FILE REFERENCE: PC-0050 US  
CURRENT APPLICATION NUMBER: US/09/907,372  
CURRENT FILING DATE: 2001-07-27  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: PERL Program  
SEQ ID NO 20  
LENGTH: 415  
TYPE: PRT  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Inocyte ID No. US20020068242A1 g600223  
US-09-907-372-20

Query Match 5.1%; Score 7; DB 9; Length 415;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 LLLAIIIL 16  
Db 222 LLLAIIIL 228

RESULT 37  
US-09-935-727-8

Sequence 8, Application US/09935727  
Patent No. US20020150583A1  
GENERAL INFORMATION:  
APPLICANT: Human Genome Sciences, Inc.  
TITLE OF INVENTION: Tumor Necrosis Factor Receptors 6 Alpha and 6 Beta  
FILE REFERENCE: PR454P2  
CURRENT APPLICATION NUMBER: US/09/935,727  
CURRENT FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: 60/303,224  
PRIOR FILING DATE: 2001-07-06  
PRIOR APPLICATION NUMBER: 60/252,131  
PRIOR FILING DATE: 2000-11-21  
PRIOR APPLICATION NUMBER: 60/227,598  
PRIOR FILING DATE: 2000-08-25

PRIOR APPLICATION NUMBER: 09/518,931  
PRIOR FILING DATE: 2000-03-03  
PRIOR APPLICATION NUMBER: 60/168,235  
PRIOR FILING DATE: 1999-12-01  
PRIOR APPLICATION NUMBER: 60/146,371  
PRIOR FILING DATE: 1999-08-02  
PRIOR APPLICATION NUMBER: 60/111,964  
PRIOR FILING DATE: 1999-04-30  
PRIOR APPLICATION NUMBER: 60/131,270  
PRIOR FILING DATE: 1999-04-27  
PRIOR APPLICATION NUMBER: 60/124,092  
PRIOR FILING DATE: 1999-03-12  
PRIOR APPLICATION NUMBER: 60/121,774  
PRIOR FILING DATE: 1999-03-04  
PRIOR APPLICATION NUMBER: 09/006,352  
PRIOR FILING DATE: 1998-01-13  
PRIOR APPLICATION NUMBER: 60/035,496  
PRIOR FILING DATE: 1997-01-14  
NUMBER OF SEQ ID NOS: 42  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 8  
LENGTH: 415  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-935-727-8

Query Match 5.1%; Score 7; DB 10; Length 415;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIIIL 16  
Db 222 LLLAIIIL 228

RESULT 38  
US-09-917-372-20  
Sequence 20, Application US/09917372  
Publication No. US20030068619A1  
GENERAL INFORMATION:  
APPLICANT: Lal, Preeti G.  
TITLE OF INVENTION: TNP RECEPTOR 2 RELATED PROTEIN VARIANT  
FILE REFERENCE: PC-0050 US  
CURRENT APPLICATION NUMBER: US/09/917,372  
CURRENT FILING DATE: 2002-09-09  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: PERL Program  
SEQ ID NO 20  
LENGTH: 415  
TYPE: PRT  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. US20030068619A1 9600223  
US-09-917-372-20

Query Match 5.1%; Score 7; DB 11; Length 415;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIIIL 16  
Db 222 LLLAIIIL 228

RESULT 39  
US-10-186-643-6  
Sequence 6, Application US/10186643  
Publication No. US20030118546A1  
GENERAL INFORMATION:  
APPLICANT: Wei, Yang-Fei  
APPLICANT: Ni, Jian

APPLICANT: Gentz, Reiner  
APPLICANT: Ruben, Steven  
TITLE OF INVENTION: Tumor Necrosis Factor Receptor 5  
FILE REFERENCE: 1488.1280004  
CURRENT APPLICATION NUMBER: US/10/186,643  
CURRENT FILING DATE: 2002-07-02  
PRIOR APPLICATION NUMBER: US/09/573,986  
PRIOR FILING DATE: 2000-05-18  
NUMBER OF SEQ ID NOS: 27  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 6  
LENGTH: 415  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-186-643-6

Query Match 5.1%; Score 7; DB 15; Length 415;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIIIL 16  
Db 222 LLLAIIIL 228

RESULT 40  
US-10-156-761-10904  
Sequence 10904, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYPEPTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 10904  
LENGTH: 417  
TYPE: PRT  
ORGANISM: Streptomyces avermitilis  
US-10-156-761-10904

Query Match 5.1%; Score 7; DB 15; Length 417;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLAIIIL 16  
Db 335 LLLAIIIL 341

RESULT 41  
US-10-098-871-8  
Sequence 8, Application US/10098871  
Publication No. US20030198958A1  
GENERAL INFORMATION:  
APPLICANT: Shimkets, Richard A.  
APPLICANT: Fernandes, Elma  
APPLICANT: Herrmann, John  
APPLICANT: Liu, Xiaohong  
APPLICANT: Yang, Meijia  
APPLICANT: Boldog, Perence  
APPLICANT: Smithson, Glendene

```

; APPLICANT: Raetelli, Luca
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND
; FILE REFERENCE: CURA-65 CIP
; CURRENT APPLICATION NUMBER: US/10/098,871
; PRIOR FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: 09/659,634
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 60/153,629
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 60/154,520
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: 60/154,762
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 60/159,231
; PRIOR FILING DATE: 2000-10-31
; PRIOR APPLICATION NUMBER: 60/276,960
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-098-871-8

Query Match
Best Local Similarity 5.1%; Score 7; DB 12; Length 429;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLAAILL 16
Db 233 LLAAILL 239

RESULT 42
US-10-108-260A-3238
; Sequence 3238, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1 full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3238
; LENGTH: 461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3238

Query Match
Best Local Similarity 5.1%; Score 7; DB 12; Length 461;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ALQALLA 13
Db 114 ALQALLA 120

RESULT 43
US-10-094-749-2349
; Sequence 2349, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
```

```

; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOMYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2349
; LENGTH: 495
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2349

Query Match
Best Local Similarity 5.1%; Score 7; DB 12; Length 495;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLAAILL 16
Db 264 LLAAILL 270

RESULT 44
US-09-769-787-80
; Sequence 80, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 80
; LENGTH: 506
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-80

Query Match
Best Local Similarity 5.1%; Score 7; DB 11; Length 506;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLAAILL 16
Db 30 LLAAILL 36

RESULT 45
US-10-114-893-10
; Sequence 10, Application US/10114893
```

Publication No. US20020193567A1  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John M.  
APPLICANT: Lavallee, Edward R.  
APPLICANT: Collins-Racie, Lisa A.  
APPLICANT: Evans, Cheryl  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Bowman, Michael R.  
APPLICANT: Spaulding, Vikki  
APPLICANT: Carlin-Duckett, McKeough  
APPLICANT: Kelleher, Kerry S.  
APPLICANT: Genetics Institute, Inc.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
FILE REFERENCE: GI 6000-10A  
CURRENT FILING DATE: 2002-04-02  
EARLIER FILING DATE: 1999-10-06  
NUMBER OF SEQ ID NOS: 321  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 10  
LENGTH: 661  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-114-893-10

Query Match 5.1%; Score 7; DB 14; Length 661;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LLLALL 16  
|||||  
Db 640 LLLALL 646

RESULT 46  
US-09-861-801-2  
Sequence 2, Application US/09861801  
Patent No. US2002009779A1  
GENERAL INFORMATION:  
APPLICANT: Meyers, Rachel A.  
APPLICANT: Williamson, Mark  
TITLE OF INVENTION: 50365, A NOVEL HEXOKINASE FAMILY MEMBER  
TITLE OF INVENTION: AND USES THEREFOR  
FILE REFERENCE: 10448-055001  
CURRENT APPLICATION NUMBER: US/09/861,801  
CURRENT FILING DATE: 2001-05-21  
PRIOR APPLICATION NUMBER: US 60/205,508  
PRIOR FILING DATE: 2000-05-19  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 917  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-861-801-2

Query Match 5.1%; Score 7; DB 9; Length 917;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 FRVLKVG 73  
|||||  
Db 90 FRVLKVG 96

RESULT 47  
US-10-170-789-28  
Sequence 28, Application US/10170789  
Publication No. US20030180930A1  
GENERAL INFORMATION:

APPLICANT: Rachel B. Meyers  
APPLICANT: Olandt, Peter J.  
APPLICANT: Kapeller-Liberman, Rosana  
APPLICANT: Curtis, Rory A. J.  
APPLICANT: Williamson, Mark  
APPLICANT: Welch, Nadine  
TITLE OF INVENTION: NOVEL HUMAN PROTEIN KINASE, PHOSPHATASE,  
AND PROTEASE FAMILY MEMBERS AND USES THEREFOR  
FILE REFERENCE: 10448-191001  
CURRENT APPLICATION NUMBER: US/10/170,789  
CURRENT FILING DATE: 2002-06-13  
PRIOR APPLICATION NUMBER: US 09/797,039  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: PCT/US01/06525  
PRIOR FILING DATE: 2001-02-28  
PRIOR APPLICATION NUMBER: US 60/186,061  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: US 09/882,166  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: PCT/US01/19269  
PRIOR FILING DATE: 2001-06-15  
PRIOR APPLICATION NUMBER: US 60/212,078  
PRIOR FILING DATE: 2000-06-15  
PRIOR APPLICATION NUMBER: US 09/934,406  
PRIOR FILING DATE: 2001-08-21  
PRIOR APPLICATION NUMBER: PCT/US01/26052  
PRIOR FILING DATE: 2001-08-21  
PRIOR APPLICATION NUMBER: US 60/226,740  
PRIOR FILING DATE: 2000-08-21  
PRIOR APPLICATION NUMBER: US 09/861,801  
PRIOR FILING DATE: 2001-05-21  
PRIOR APPLICATION NUMBER: PCT/US01/16549  
PRIOR FILING DATE: 2001-05-21  
PRIOR APPLICATION NUMBER: US 60/205,508  
PRIOR FILING DATE: 2000-05-19  
PRIOR APPLICATION NUMBER: US 09/801,267  
PRIOR FILING DATE: 2001-03-06  
PRIOR APPLICATION NUMBER: PCT/US01/07138  
PRIOR FILING DATE: 2001-03-05  
PRIOR APPLICATION NUMBER: US 60/187,454  
PRIOR FILING DATE: 2000-03-07  
PRIOR APPLICATION NUMBER: US 09/829,671  
PRIOR FILING DATE: 2001-04-10  
PRIOR APPLICATION NUMBER: PCT/US01/40483  
PRIOR FILING DATE: 2001-04-11  
PRIOR APPLICATION NUMBER: US 60/197,508  
PRIOR FILING DATE: 2000-04-18  
PRIOR APPLICATION NUMBER: US 09/961,721  
PRIOR FILING DATE: 2001-09-24  
PRIOR APPLICATION NUMBER: PCT/US01/29904  
PRIOR FILING DATE: 2001-09-24  
PRIOR APPLICATION NUMBER: US 60/235,023  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: US 10/045,367  
PRIOR FILING DATE: 2001-11-07  
PRIOR APPLICATION NUMBER: US 60/246,561  
PRIOR FILING DATE: 2000-11-07  
PRIOR APPLICATION NUMBER: US 09/801,275  
PRIOR FILING DATE: 2001-03-06  
PRIOR APPLICATION NUMBER: PCT/US01/07074  
PRIOR FILING DATE: 2001-03-05  
PRIOR APPLICATION NUMBER: US 60/187,420  
PRIOR FILING DATE: 2000-03-07  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 28  
LENGTH: 917  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-170-789-28

Query Match 5.1%; Score 7; DB 12; Length 917;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 FRVLKQV 73  
 |||||  
 Db 90 FRVLKQV 96

RESULT 48  
 US-10-108-260A-2966  
 ; Sequence 2966, Application US/10108260A  
 ; Publication No. US20040005560A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HELIX RESEARCH INSTITUTE  
 ; TITLE OF INVENTION: No. US20040005560A1 full length cDNA  
 ; FILE REFERENCE: H1-A0106  
 ; CURRENT APPLICATION NUMBER: US/10/108,260A  
 ; CURRENT FILING DATE: 2002-03-27  
 ; NUMBER OF SEQ ID NOS: 5458  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2966  
 ; LENGTH: 917  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-108-260A-2966

Query Match 5.1%; Score 7; DB 12; Length 917;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 FRVLKQV 73  
 |||||  
 Db 90 FRVLKQV 96

RESULT 49  
 US-10-262-794A-53  
 ; Sequence 53, Application US/10262794A  
 ; Publication No. US20030207806A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ensign, Jerald C  
 ; APPLICANT: Bowen, David J  
 ; APPLICANT: Petell, James  
 ; APPLICANT: Fatig, Raymond  
 ; APPLICANT: Schoonover, Sue  
 ; APPLICANT: French-Constant, Richard  
 ; APPLICANT: Rocheleau, Thomas A.  
 ; APPLICANT: Blackburn, Michael B.  
 ; APPLICANT: Hey, Timothy D.  
 ; APPLICANT: Merlo, Donald J.  
 ; APPLICANT: Orr, Gregory L.  
 ; APPLICANT: Roberts, Jean L.  
 ; APPLICANT: Strickland, James A.  
 ; APPLICANT: Guo, Lining  
 ; APPLICANT: Cliche, Todd A.  
 ; APPLICANT: Sukhaphinda, Kitisri  
 ; TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus  
 ; NUMBER OF SEQUENCES: 88  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Dow Agrosciences Patent Department  
 ; STREET: 9330 Zionsville Road  
 ; CITY: Indianapolis  
 ; STATE: IN  
 ; COUNTRY: US  
 ; ZIP: 46268  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/10/262,794A  
 ; FILING DATE: 02-OCT-2002  
 ; CLASSIFICATION:

PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/851,567  
 ; FILING DATE: 05-MAY-1997  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/063,615  
 ; FILING DATE: 18-MAY-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/395,497  
 ; FILING DATE: 28-FEB-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/007,255  
 ; FILING DATE: 06-NOV-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/608,423  
 ; FILING DATE: 28-FEB-1996  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/705,484  
 ; FILING DATE: 28-AUG-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Seay, Nicholas J  
 ; REGISTRATION NUMBER: 27386  
 ; REFERENCE/DOCKET NUMBER: 960296.93804  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 608-251-5000  
 ; TELEFAX: 608-251-9166  
 ; INFORMATION FOR SEQ ID NO: 53:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1844 amino acids  
 ; TYPE: amino acids  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-10-262-794A-53

Query Match 5.1%; Score 7; DB 12; Length 1844;  
 Best Local Similarity 100.0%; Pred. No. 5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 LITLMAL 21  
 |||||  
 Db 712 LITLMAL 718

RESULT 50  
 US-10-242-056-53  
 ; Sequence 53, Application US/10242056  
 ; Publication No. US20030113323A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ensign, Jerald C  
 ; APPLICANT: Bowen, David J  
 ; APPLICANT: Petell, James  
 ; APPLICANT: Fatig, Raymond  
 ; APPLICANT: Schoonover, Sue  
 ; APPLICANT: French-Constant, Richard  
 ; APPLICANT: Merlo, Donald J.  
 ; APPLICANT: Orr, Gregory L.  
 ; APPLICANT: Roberts, Jean L.  
 ; APPLICANT: Rocheleau, Thomas A.  
 ; TITLE OF INVENTION: Insecticidal Protein Toxins from  
 ; NUMBER OF SEQUENCES: 88  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: DowBianco  
 ; STREET: 9330 Zionsville Road  
 ; CITY: Indianapolis  
 ; STATE: IN  
 ; COUNTRY: US  
 ; ZIP: 46268  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/242,056  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/063,615  
FILING DATE: 18-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/395,497  
FILING DATE: 28-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/007,255  
FILING DATE: 06-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/608,423  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/705,484  
FILING DATE: 28-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/743,699  
FILING DATE: 06-NOV-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Botucki, Andrea T.  
REGISTRATION NUMBER: 33651  
REFERENCE/DOCKET NUMBER: 50301E  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-337-4846  
TELEFAX: 317-337-4847  
INFORMATION FOR SEQ. ID NO: 53:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1844 amino acids  
TYPE: amino acids  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-10-242-056-53

Query Match 5.1%; Score 7; DB 15; Length 1844;  
Best Local Similarity 100.0%; Pred. No. 5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 15 LITMAL 21  
Db 712 LITMAL 718

Search completed: January 21, 2004, 12:33:00  
Job time : 34 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: January 21, 2004, 12:00:29 ; Search time 65.4064 Seconds  
(without alignments)  
332.468 Million cell updates/sec

Title: US-09-941-314-2

Sequence: 1 MABPWQALQLLAILLTLM.....VPAVPWFQYKTLNKSCSSD 137

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

A\_Geneseq\_19Jun03:\*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	736	100.0	137	23	AAU79852	Human cystatin-8
2	640	87.0	117	23	AAU79854	Human cystatin-8
3	629	85.5	115	23	AAU79853	Human cystatin-8
4	446	60.6	80	23	AAU79865	Human cystatin-8
5	334	45.4	59	23	AAU79866	Human cystatin-8
6	288	39.1	52	23	AAU79864	Human cystatin-8
7	273	37.1	48	23	AAU79867	Human cystatin-8
8	272	37.0	49	23	AAU79863	Human cystatin-8
9	262	35.6	50	22	ABG48915	Human liver peptid

10	262	35.6	50	22	ABR28913	Peptide #1564 enco
11	262	35.6	50	22	ABR32389	Peptide #5040 enco
12	262	35.6	50	22	ABR34086	Peptide #1592 enco
13	262	35.6	50	22	ABR19524	Protein #1523 enco
14	262	35.6	50	22	AAW54871	Human brain expro
15	262	35.6	50	22	AAW67252	Human bone marrow
16	262	35.6	50	22	AAW15096	Peptide #1530 enco
17	262	35.6	50	22	AAW27545	Peptide #1582 enco
18	262	35.6	50	22	AAW02833	Peptide #1515 enco
19	262	35.6	50	23	ABG36903	Human peptid enco
20	254	34.5	46	23	AAW04433	Human cystatin-8 (
21	226.5	30.8	142	22	AAW02404	Mouse cystatin-rel
22	226.5	30.8	142	22	AAW78258	Fragment of human
23	202	27.4	142	20	AAW04434	Human cystatin-rel
24	202	27.4	142	22	AAW02405	Human cystatin-rel
25	202	27.4	142	23	ABG60085	Human DTHP polype
26	201.5	27.4	141	21	AAW96576	Murine cystatin T
27	195.5	26.6	141	21	AAW04432	Mouse testis spect
28	195.5	26.6	141	22	AAW02403	Murine cystatin T
29	195.5	26.6	141	22	AAW79862	Human cystatin-8 (
30	189	25.7	33	23	AAU79856	Fragment of human
31	189	25.7	36	23	AAW78260	Human cystatin-8 (
32	188.5	25.6	123	20	AAW79858	Human cystatin-8 (
33	187	25.4	35	23	AAW53770	Amino acid sequenc
34	186.5	25.3	140	23	ABR40483	Human secreted pro
35	186.5	25.3	140	23	ABR40484	Human secreted pro
36	186.5	25.3	140	22	AAW04315	Alternative versio
37	186.5	25.3	145	22	AAW76555	Human Zcy5 polype
38	186.5	25.3	145	23	ABG75917	Human cystatin sup
39	186.5	25.3	92	20	AAW78259	Fragment of human
40	184.5	25.1	140	23	ABR40411	Human secreted pro
41	180.5	24.5	140	22	AAW04333	Human Zcy5 protei
42	180.5	24.5	145	22	AAW04887	Human protease pro
43	180.5	24.5	145	23	AAW76578	Human Zcy5 polype
44	180.5	24.5	145	23	ABG75925	Human cystatin sup
45	180.5	24.5	145	24		

## ALIGNMENTS

RESULT 1	AAU79852	standard; Protein; 137 AA.
ID	AAU79852	
XX	AAU79852;	
AC	15-JUL-2002 (first entry)	
XX		
DT	Human cystatin-8 (Zcy58)	
XX		
DE	Cystatin-8; Zcy58; cancer; procoagulant protein; thrombosis;	
XX	Human sperm motility; fertilisation.	
KW	Homo sapiens.	
XX	WO200220567-A2.	
XX	14-MAR-2002.	
PD	29-AUG-2001; 2001MO-US26868.	
XX	01-SEP-2000; 2000US-230230P.	
PR	(ZYMO) ZYMOGENETICS INC.	
PA	Holloway JL, Gao Z, Bishop PD, et al	
XX	WPI; 2002-383044/41.	
DR	N-PSDB; ABK49522.	
XX	Novel isolated mammalian cystatin-8 polypeptide useful for promoting	

PT spermatoogenesis, and inhibiting cancer procoagulant protein which leads  
 PT to inhibition of thrombotic events associated with cancer -  
 PS Claim 2, Page 93-94, 100pp, English.

XX The invention describes an isolated mammalian cystatin-8 (Zcys8)  
 CC polypeptide (I). (I) is useful for: inhibiting cancer procoagulant  
 CC protein in an individual and thus inhibiting the thrombotic events  
 CC associated with cancer; promoting spermatoogenesis, modulating seminal  
 CC fluid viscosity, enhancing viability of cryopreserved sperm, sperm  
 CC motility and fertilisation; and as antigenic peptides to generate  
 CC antibodies. Zcys8 is useful as research reagent for characterising sites  
 CC of interaction between Zcys8 and its receptor. Zcys8 is useful in  
 CC enhancing fertilisation during assisted reproduction in humans and in  
 CC animals. Anti-(I) antibodies are useful to screen biological samples like  
 CC blood, urine, saliva, tissue biopsy and autopsy material in vitro for the  
 CC presence of Zcys8. The antibodies are also useful to isolate large  
 CC quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.  
 CC The polynucleotide encoding (I) is useful to detect and to localise the  
 CC expression of a Zcys8 gene in a biological sample and Zcys8  
 CC oligonucleotide probes are useful for in vivo diagnosis. The  
 CC polynucleotide encoding (I) is useful in determining whether a subject's  
 CC chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene  
 CC copy number changes, insertions, deletions, restriction site changes and  
 CC rearrangements and genetic alterations that inactivate the Zcys8 gene.  
 CC This is the amino acid sequence of human cystatin-8 (Zcys8).

XX Sequence 137 AA;

Query Match 100.0%; Score 736; DB 23; Length 137;  
 Best Local Similarity 100.0%; Pred. No. 7, 4e-75;  
 Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAEFWQALQILLALITLALPYQARKKFLSVHEVNAVENYAKDSLOWITDQNKESDD 60  
 DB 1 MAEFWQALQILLALITLALPYQARKKFLSVHEVNAVENYAKDSLOWITDQNKESDD 60  
 QY 61 KYHRRIRRVKVGQVVDHLEHYNVEMQWTTCKPEPTTNCVPERELHKQVCFPSVFA 120  
 DB 61 KYHRRIRRVKVGQVVDHLEHYNVEMQWTTCKPEPTTNCVPERELHKQVCFPSVFA 120  
 QY 121 VPWFQYKIINKSCSSD 137  
 DB 121 VPWFQYKIINKSCSSD 137

RESULT 2  
 AAU79854  
 ID AAU79854 standard; Protein; 117 AA.

XX AC AAU79854;  
 XX DT 15-JUL-2002 (first entry)  
 XX DE Human cystatin-8 (Zcys8) antigenic fragment #2.  
 XX KW Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;  
 XX KM spermatoogenesis; seminal fluid viscosity; cryopreserved sperm;  
 XX KW sperm motility; fertilisation; antigenic fragment.  
 XX OS Homo sapiens.  
 XX PN WO200220567-A2.  
 XX PD 14-MAR-2002.  
 XX PF 29-AUG-2001; 2001WO-US26868.  
 XX PR 01-SEP-2000; 2000US-230230P.  
 XX PA (ZYMO) ZYMOGENETICS INC.  
 XX PI Holloway JL, Gao Z, Bishop PD;

XX WPI; 2002-383044/41.

DR Novel isolated mammalian cystatin-8 polypeptide useful for promoting  
 XX spermatoogenesis, and inhibiting cancer procoagulant protein which leads  
 PT to inhibition of thrombotic events associated with cancer -  
 PS Claim 2, Page 94-95, 100pp, English.

XX The invention describes an isolated mammalian cystatin-8 (Zcys8)  
 CC polypeptide (I). (I) is useful for: inhibiting cancer procoagulant  
 CC protein in an individual and thus inhibiting the thrombotic events  
 CC associated with cancer; promoting spermatoogenesis, modulating seminal  
 CC fluid viscosity, enhancing viability of cryopreserved sperm, sperm  
 CC motility and fertilisation; and as antigenic peptides to generate  
 CC antibodies. Zcys8 is useful as research reagent for characterising sites  
 CC of interaction between Zcys8 and its receptor. Zcys8 is useful in  
 CC enhancing fertilisation during assisted reproduction in humans and in  
 CC animals. Anti-(I) antibodies are useful to screen biological samples like  
 CC blood, urine, saliva, tissue biopsy and autopsy material in vitro for the  
 CC presence of Zcys8. The antibodies are also useful to isolate large  
 CC quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.  
 CC The polynucleotide encoding (I) is useful to detect and to localise the  
 CC expression of a Zcys8 gene in a biological sample and Zcys8  
 CC oligonucleotide probes are useful for in vivo diagnosis. The  
 CC polynucleotide encoding (I) is useful in determining whether a subject's  
 CC chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene  
 CC copy number changes, insertions, deletions, restriction site changes and  
 CC rearrangements and genetic alterations that inactivate the Zcys8 gene.  
 CC This sequence represents an antigenic fragment of human cystatin-8  
 CC (Zcys8).

XX Sequence 117 AA;

Query Match 87.0%; Score 640; DB 23; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 3, 9e-64;  
 Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 LPYQARKKFLSVHEVNAVENYAKDSLOWITDQNKESDDKYHRRIRRVKVGQVVDHL 80  
 DB 1 LPYQARKKFLSVHEVNAVENYAKDSLOWITDQNKESDDKYHRRIRRVKVGQVVDHL 80  
 QY 81 EYHLNVEMQWTTCKPEPTTNCVPERELHKQVCFPSVFAVPWFQYKIINKSCSSD 137  
 DB 61 EYHLNVEMQWTTCKPEPTTNCVPERELHKQVCFPSVFAVPWFQYKIINKSCSSD 117

RESULT 3  
 AAU79853  
 ID AAU79853 standard; Protein; 115 AA.

XX AC AAU79853;  
 XX DT 15-JUL-2002 (first entry)  
 XX DE Human cystatin-8 (Zcys8) antigenic fragment #1.  
 XX KW Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;  
 XX KM spermatoogenesis; seminal fluid viscosity; cryopreserved sperm;  
 XX KW sperm motility; fertilisation; antigenic fragment.  
 XX OS Homo sapiens.  
 XX PN WO200220567-A2.  
 XX PD 14-MAR-2002.  
 XX PF 29-AUG-2001; 2001WO-US26868.  
 XX PR 01-SEP-2000; 2000US-230230P.  
 XX PA (ZYMO) ZYMOGENETICS INC.

PI Holloway JL, Gao Z, Bishop PD;  
 XX  
 XX WPI; 2002-383044/41.  
 DR  
 XX Novel isolated mammalian cystatin-8 polypeptide useful for promoting  
 PT spermatogenesis, and inhibiting cancer procoagulant protein which leads  
 PT to inhibition of thrombotic events associated with cancer -  
 XX  
 XX Claim 2; Page 94; 100pp; English.  
 PS  
 XX The invention describes an isolated mammalian cystatin-8 (Zcys8)  
 CC polypeptide (I). (I) is useful for: inhibiting cancer procoagulant  
 CC protein in an individual and thus inhibiting the thrombotic events  
 CC associated with cancer; promoting spermatogenesis, modulating seminal  
 CC fluid viscosity, enhancing viability of cryopreserved sperm, sperm  
 CC motility and fertilisation; and as antigenic peptides to generate  
 CC antibodies. Zcys8 is useful as research reagent for characterising sites  
 CC of interaction between Zcys8 and its receptor. Zcys8 is useful in  
 CC enhancing fertilisation during assisted reproduction in humans and in  
 CC animals. Anti-(I) antibodies are useful to screen biological samples like  
 CC blood, urine, saliva, tissue biopsy and autopsy material in vitro for the  
 CC presence of Zcys8. The antibodies are also useful to isolate large  
 CC quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.  
 CC The polynucleotide encoding (I) is useful to detect and to localise the  
 CC expression of a Zcys8 gene in a biological sample and Zcys8  
 CC oligonucleotide probes are useful for in vivo diagnosis. The  
 CC polynucleotide encoding (I) is useful in determining whether a subject's  
 CC chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene  
 CC copy number changes, insertions, deletions, restriction site changes and  
 CC rearrangements and genetic alterations that inactivate the Zcys8 gene.  
 CC This sequence represents an antigenic fragment of human cystatin-8  
 CC (Zcys8).  
 CC  
 SQ Sequence 115 AA;  
 XX  
 XX Query Match 85.5%; Score 629; DB 23; Length 115;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-63;  
 Matches 115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 23 YQARKTFLSVHEVMAVENYAKDSLQWITDQYNKSDDKYHFRFLVKYQROVTDHLEY 82  
 DB 1 YQARKTFLSVHEVMAVENYAKDSLQWITDQYNKSDDKYHFRFLVKYQROVTDHLEY 60  
 QY 83 HLNVENQMTTCQRPETNNCPQREELHKOYNCFFSVFAVWPFQYKILNKSCSSD 137  
 DB 61 HLNVENQMTTCQRPETNNCPQREELHKOYNCFFSVFAVWPFQYKILNKSCSSD 115  
 RESULT 4  
 AAU79865  
 ID AAU79865 standard; Peptide; 80 AA.  
 XX  
 AC AAU79865;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Human cystatin-8 (Zcys8) antigenic fragment #13.  
 XX  
 KM Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;  
 KW spermatogenesis; seminal fluid viscosity; cryopreserved sperm;  
 KM sperm motility; fertilisation; antigenic peptide.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200220567-A2.  
 XX  
 PD 14-MAR-2002.  
 XX  
 PF 29-AUG-2001; 2001WO-US26868.  
 XX  
 PR 01-SEP-2000; 2000US-230230P.  
 XX  
 PA (ZYMO ) ZYMOGENETICS INC.

XX  
 PI Holloway JL, Gao Z, Bishop PD;  
 XX  
 XX WPI; 2002-383044/41.  
 DR  
 XX Novel isolated mammalian cystatin-8 polypeptide useful for promoting  
 PT spermatogenesis, and inhibiting cancer procoagulant protein which leads  
 PT to inhibition of thrombotic events associated with cancer -  
 XX  
 XX Claim 2; Page 98; 100pp; English.  
 PS  
 XX The invention describes an isolated mammalian cystatin-8 (Zcys8)  
 CC polypeptide (I). (I) is useful for: inhibiting cancer procoagulant  
 CC protein in an individual and thus inhibiting the thrombotic events  
 CC associated with cancer; promoting spermatogenesis, modulating seminal  
 CC fluid viscosity, enhancing viability of cryopreserved sperm, sperm  
 CC motility and fertilisation; and as antigenic peptides to generate  
 CC antibodies. Zcys8 is useful as research reagent for characterising sites  
 CC of interaction between Zcys8 and its receptor. Zcys8 is useful in  
 CC enhancing fertilisation during assisted reproduction in humans and in  
 CC animals. Anti-(I) antibodies are useful to screen biological samples like  
 CC blood, urine, saliva, tissue biopsy and autopsy material in vitro for the  
 CC presence of Zcys8. The antibodies are also useful to isolate large  
 CC quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.  
 CC The polynucleotide encoding (I) is useful to detect and to localise the  
 CC expression of a Zcys8 gene in a biological sample and Zcys8  
 CC oligonucleotide probes are useful for in vivo diagnosis. The  
 CC polynucleotide encoding (I) is useful in determining whether a subject's  
 CC chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene  
 CC copy number changes, insertions, deletions, restriction site changes and  
 CC rearrangements and genetic alterations that inactivate the Zcys8 gene.  
 CC This sequence represents an antigenic peptide of human cystatin-8  
 CC (Zcys8).  
 CC  
 SQ Sequence 80 AA;  
 XX  
 XX Query Match 60.6%; Score 446; DB 23; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-42;  
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 53 QYNKESDDKYNFRIFRVLKQROVTDHLEYHLNVENQMTTCQRPETNNCPQREELHKOY 112  
 DB 1 QYNKESDDKYNFRIFRVLKQROVTDHLEYHLNVENQMTTCQRPETNNCPQREELHKOY 60  
 QY 113 NCFPSVFAVWPFQYKILNK 132  
 DB 61 NCFPSVFAVWPFQYKILNK 80  
 RESULT 5  
 AAU79866  
 ID AAU79866 standard; Peptide; 59 AA.  
 XX  
 AC AAU79866;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Human cystatin-8 (Zcys8) antigenic fragment #14.  
 XX  
 KM Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;  
 KW spermatogenesis; seminal fluid viscosity; cryopreserved sperm;  
 KM sperm motility; fertilisation; antigenic peptide.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200220567-A2.  
 XX  
 PD 14-MAR-2002.  
 XX  
 PF 29-AUG-2001; 2001WO-US26868.  
 XX  
 PR 01-SEP-2000; 2000US-230230P.  
 XX

PA (ZYMO ) ZYMOGENETICS INC.  
XX  
PI Holloway JL, Gao Z, Bishop PD;  
XX  
XX WPI; 2002-383044/41.  
XX  
PT Novel isolated mammalian cystatin-8 polypeptide useful for promoting  
XX spermatogenesis, and inhibiting cancer procoagulant protein which leads  
XX to inhibition of thrombotic events associated with cancer  
XX  
PS Claim 2; Page 99; 100pp; English.  
XX  
XX The invention describes an isolated mammalian cystatin-8 (Zcys8)  
XX polypeptide (I). (I) is useful for: inhibiting cancer procoagulant  
XX protein in an individual and thus inhibiting the thrombotic events  
XX associated with cancer; promoting spermatogenesis, modulating seminal  
XX fluid viscosity, enhancing viability of cryopreserved sperm, sperm  
XX motility and fertilisation; and as antigenic peptides to generate  
XX antibodies. Zcys8 is useful as research reagent for characterising sites  
XX of interaction between Zcys8 and its receptor. Zcys8 is useful in  
XX enhancing fertilisation during assisted reproduction in humans and in  
XX animals. Anti-(I) antibodies are useful to screen biological samples like  
XX blood, urine, saliva, tissue biopsy and autopsy material in vitro for the  
XX presence of Zcys8. The antibodies are also useful to isolate large  
XX quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.  
XX The polynucleotide encoding (I) is useful to detect and to localise the  
XX expression of a Zcys8 gene in a biological sample and Zcys8  
XX oligonucleotide probes are useful for in vivo diagnosis. The  
XX polynucleotide encoding (I) is useful in determining whether a subject's  
XX chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene  
XX copy number changes, insertions, deletions, restriction site changes and  
XX rearrangements and genetic alterations that inactivate the Zcys8 gene.  
XX This sequence represents an antigenic peptide of human cystatin-8  
XX (Zcys8).  
XX  
SQ Sequence 59 AA;  
XX  
XX  
Query Match 45.4%; Score 334; DB 23; Length 59;  
Best Local Similarity 100.0%; Pred. No. 4.7e-30;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 74 RQVTDHLEHNLVNMOWTTCQKPEPTNCVPORELIHKVNCFPFVPAVPMFEQYKILNK 132  
Db 1 RQVTDHLEHNLVNMOWTTCQKPEPTNCVPORELIHKVNCFPFVPAVPMFEQYKILNK 59  
RESULT 6  
AAU79864  
ID AAU79864 standard; Peptide; 52 AA.  
XX  
AC AAU79864;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Human cystatin-8 (Zcys8) antigenic fragment #12.  
XX  
XX Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;  
XX spermatogenesis; seminal fluid viscosity; cryopreserved sperm;  
XX sperm motility; fertilisation; antigenic peptide.  
XX  
OS Homo sapiens.  
XX  
XX WO200220567-A2.  
XX  
XX 14-MAR-2002.  
XX  
XX 29-AUG-2001; 2001WO-US26868.  
XX  
XX 01-SEP-2000; 2000US-230230P.  
XX  
XX (ZYMO ) ZYMOGENETICS INC.  
XX  
XX Holloway JL, Gao Z, Bishop PD;  
XX  
XX

XX  
XX WPI; 2002-383044/41.  
XX  
XX Novel isolated mammalian cystatin-8 polypeptide useful for promoting  
XX spermatogenesis, and inhibiting cancer procoagulant protein which leads  
XX to inhibition of thrombotic events associated with cancer  
XX  
PS Claim 2; Page 98; 100pp; English.  
XX  
XX The invention describes an isolated mammalian cystatin-8 (Zcys8)  
XX polypeptide (I). (I) is useful for: inhibiting cancer procoagulant  
XX protein in an individual and thus inhibiting the thrombotic events  
XX associated with cancer; promoting spermatogenesis, modulating seminal  
XX fluid viscosity, enhancing viability of cryopreserved sperm, sperm  
XX motility and fertilisation; and as antigenic peptides to generate  
XX antibodies. Zcys8 is useful as research reagent for characterising sites  
XX of interaction between Zcys8 and its receptor. Zcys8 is useful in  
XX enhancing fertilisation during assisted reproduction in humans and in  
XX animals. Anti-(I) antibodies are useful to screen biological samples like  
XX blood, urine, saliva, tissue biopsy and autopsy material in vitro for the  
XX presence of Zcys8. The antibodies are also useful to isolate large  
XX quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.  
XX The polynucleotide encoding (I) is useful to detect and to localise the  
XX expression of a Zcys8 gene in a biological sample and Zcys8  
XX oligonucleotide probes are useful for in vivo diagnosis. The  
XX polynucleotide encoding (I) is useful in determining whether a subject's  
XX chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene  
XX copy number changes, insertions, deletions, restriction site changes and  
XX rearrangements and genetic alterations that inactivate the Zcys8 gene.  
XX This sequence represents an antigenic peptide of human cystatin-8  
XX (Zcys8).  
XX  
SQ Sequence 52 AA;  
XX  
XX  
Query Match 39.1%; Score 288; DB 23; Length 52;  
Best Local Similarity 100.0%; Pred. No. 6.1e-25;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 56 KESDDKXHFRIFFVLKYQROVTDHLEHNLVNMOWTTCQKPEPTNCVPORE 107  
Db 1 KESDDKXHFRIFFVLKYQROVTDHLEHNLVNMOWTTCQKPEPTNCVPORE 52  
RESULT 7  
AAU79867  
ID AAU79867 standard; Peptide; 48 AA.  
XX  
AC AAU79867;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Human cystatin-8 (Zcys8) antigenic fragment #15.  
XX  
XX Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;  
XX spermatogenesis; seminal fluid viscosity; cryopreserved sperm;  
XX sperm motility; fertilisation; antigenic peptide.  
XX  
OS Homo sapiens.  
XX  
XX WO200220567-A2.  
XX  
XX 14-MAR-2002.  
XX  
XX 29-AUG-2001; 2001WO-US26868.  
XX  
XX 01-SEP-2000; 2000US-230230P.  
XX  
XX (ZYMO ) ZYMOGENETICS INC.  
XX  
XX Holloway JL, Gao Z, Bishop PD;  
XX  
XX WPI; 2002-383044/41.  
XX  
XX

PT Novel isolated mammalian cystatin-8 polypeptide useful for promoting  
 PT spermatogenesis, and inhibiting cancer procoagulant protein which leads  
 PT to inhibition of thrombotic events associated with cancer  
 XX  
 XX Claim 2; Page 99; 100pp; English.  
 PS  
 CC The invention describes an isolated mammalian cystatin-8 (Zcys8)  
 CC polypeptide (1). (1) is useful for: inhibiting cancer procoagulant  
 CC protein in an individual and thus inhibiting the thrombotic events  
 CC associated with cancer; promoting spermatogenesis, modulating seminal  
 CC fluid viscosity, enhancing viability of cryopreserved sperm, sperm  
 CC motility and fertilisation; and as antigenic peptides to generate  
 CC antibodies. Zcys8 is useful as research reagent for characterising sites  
 CC of interaction between Zcys8 and its receptor. Zcys8 is useful in  
 CC enhancing fertilisation during assisted reproduction in humans and in  
 CC animals. Anti-(1) antibodies are useful to screen biological samples like  
 CC blood, urine, saliva, tissue biopsy and autopsy material in vitro for the  
 CC presence of Zcys8. The antibodies are also useful to isolate large  
 CC quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.  
 CC The polynucleotide encoding (1) is useful to detect and to localise the  
 CC expression of a Zcys8 gene in a biological sample and Zcys8  
 CC oligonucleotide probes are useful for in vivo diagnosis. The  
 CC polynucleotide encoding (1) is useful in determining whether a subject's  
 CC chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene  
 CC copy number changes, insertions, deletions, restriction site changes and  
 CC rearrangements and genetic alterations that inactivate the Zcys8 gene.  
 CC This sequence represents an antigenic peptide of human cystatin-8  
 CC (Zcys8).  
 CC  
 CC Sequence 48 AA;  
 CC  
 SQ  
 Query Match 37.1%; Score 273; DB 23; Length 48;  
 Best Local Similarity 100.0%; Pred. No. 2.7e-23;  
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 85 NVNEMQWTTCKPRTTNCVPERELHKQVNCFSYFAVPMFPOYKILNK 132  
 Db 1 NVNEMQWTTCKPRTTNCVPERELHKQVNCFSYFAVPMFPOYKILNK 48  
 RESULT 8  
 AAU79863  
 ID AAU79863 standard; Peptide; 49 AA.  
 AC AAU79863;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Human cystatin-8 (Zcys8) antigenic fragment #11.  
 XX  
 KM Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;  
 KM spermatogenesis; seminal fluid viscosity; cryopreserved sperm;  
 KM sperm motility; fertilisation; antigenic peptide.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO200220567-A2.  
 XX  
 PD 14-MAR-2002.  
 XX  
 PF 29-AUG-2001; 2001WO-US26868.  
 XX  
 PR 01-SEP-2000; 2000US-230230P.  
 XX  
 PR (ZYMO) ZYMOGENETICS INC.  
 PA  
 PI Holloway JL, Gao Z, Bishop PD;  
 XX  
 DR WPI; 2002-383044/41.  
 XX  
 PT Novel isolated mammalian cystatin-8 polypeptide useful for promoting  
 PT spermatogenesis, and inhibiting cancer procoagulant protein which leads  
 PT to inhibition of thrombotic events associated with cancer

XX  
 PS Claim 2; Page 97-98; 100pp; English.  
 XX  
 CC The invention describes an isolated mammalian cystatin-8 (Zcys8)  
 CC polypeptide (1). (1) is useful for: inhibiting cancer procoagulant  
 CC protein in an individual and thus inhibiting the thrombotic events  
 CC associated with cancer; promoting spermatogenesis, modulating seminal  
 CC fluid viscosity, enhancing viability of cryopreserved sperm, sperm  
 CC motility and fertilisation; and as antigenic peptides to generate  
 CC antibodies. Zcys8 is useful as research reagent for characterising sites  
 CC of interaction between Zcys8 and its receptor. Zcys8 is useful in  
 CC enhancing fertilisation during assisted reproduction in humans and in  
 CC animals. Anti-(1) antibodies are useful to screen biological samples like  
 CC blood, urine, saliva, tissue biopsy and autopsy material in vitro for the  
 CC presence of Zcys8. The antibodies are also useful to isolate large  
 CC quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.  
 CC The polynucleotide encoding (1) is useful to detect and to localise the  
 CC expression of a Zcys8 gene in a biological sample and Zcys8  
 CC oligonucleotide probes are useful for in vivo diagnosis. The  
 CC polynucleotide encoding (1) is useful in determining whether a subject's  
 CC chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene  
 CC copy number changes, insertions, deletions, restriction site changes and  
 CC rearrangements and genetic alterations that inactivate the Zcys8 gene.  
 CC This sequence represents an antigenic peptide of human cystatin-8  
 CC (Zcys8).  
 CC  
 CC Sequence 49 AA;  
 CC  
 SQ  
 Query Match 37.0%; Score 272; DB 23; Length 49;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-23;  
 Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 52 DQYNKSSDDKXFRIRFVILKVRQVTDHLEYNVEMQWTTCKPRTTN 100  
 Db 1 DQYNKSSDDKXFRIRFVILKVRQVTDHLEYNVEMQWTTCKPRTTN 49  
 RESULT 9  
 ABG48915  
 ID ABG48915 standard; Peptide; 50 AA.  
 XX  
 AC ABG48915;  
 XX  
 DT 25-FEB-2003 (first entry)  
 XX  
 DE Human liver peptide, SEQ ID No 27563.  
 XX  
 KM Human; liver; cirrhosis; hyperlipoproteinemia; hyperlipidaemia;  
 KM hypercholesterolaemia; coronary heart disease.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO200157273-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00664.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLFE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-488998/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for

PT analysing gene expression in human adult liver -  
 XX  
 PS Claim 27; SEQ ID No 27563; 658bp; English.  
 XX  
 CC The invention relates to a single exon nucleic acid probe (SENP) (1) for  
 CC measuring human gene expression in a sample derived from human adult  
 CC liver, comprising one of 13109 defined nucleotide sequences given in the  
 CC specification (or complements/ fragments). The probe hybridises at high  
 CC stringency to a nucleic acid molecule expressed in the human adult  
 CC liver. (1) may be used for predicting, measuring and displaying gene  
 CC expression in samples derived from human adult liver. The genes  
 CC identified may be involved in genetic liver diseases such as cirrhosis,  
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which  
 CC is associated with coronary heart disease. AB647348-AB65930 represent  
 CC human liver single exon encoded peptides of the invention.  
 CC Note: The sequence information for this patent does not appear in the  
 CC printed specification but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 50 AA;  
 Query Match 35.6%; Score 262; DB 22; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 4.9e-22;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 26 RKKTFLSVHEWMAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKYQQR 75  
 Db 1 RKKTFLSVHEWMAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKYQQR 50  
 RESULT 10  
 ABB28913  
 ID ABB28913 standard; Peptide; 50 AA.  
 XX  
 AC ABB28913;  
 XX  
 DT 01-FEB-2002 (first entry)  
 XX  
 DE Peptide #1564 encoded by breast cell single exon nucleic acid probe.  
 XX  
 KW Human; microarray; single exon probe; gene expression; breast;  
 KW disease; cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157271-A2.  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US000662.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 DR WPI; 2001-496933/54.  
 XX  
 PT New spatially-addressable set of single exon nucleic acid probes,  
 PT useful for measuring gene expression in sample derived from human  
 PT breast, comprises number of single exon nucleic acid probes -  
 XX  
 PS Claim 27; SEQ ID NO 11881; 327bp + sequence listing; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived

CC from human breast and BT 474 cells. The method involves contacting  
 CC the probes with a collection of detectably labelled nucleic acids  
 CC derived from mRNA of human breast, and then measuring the label  
 CC bound to each probe of the microarray. The probes are useful for  
 CC verifying the expression of regions of genomic DNA predicted to  
 CC encode proteins. They are useful for gene discovery, and for  
 CC determining predisposition and/or prognosing breast disease. Gene  
 CC expression analysis is useful for assessing the toxicity of chemical  
 CC agents on cells. The microarray of this invention presents a far greater  
 CC diversity of probes for measuring gene expression, with far less bias  
 CC than expressed sequence tag microarrays. The method is suitable for  
 CC rapid production of functional information from genomic sequence. The  
 CC present sequence is a peptide encoded by a single exon nucleic acid  
 CC probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 50 AA;  
 Query Match 35.6%; Score 262; DB 22; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 4.9e-22;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 26 RKKTFLSVHEWMAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKYQQR 75  
 Db 1 RKKTFLSVHEWMAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKYQQR 50  
 RESULT 11  
 ABB32389  
 ID ABB32389 standard; Peptide; 50 AA.  
 XX  
 AC ABB32389;  
 XX  
 DT 01-FEB-2002 (first entry)  
 XX  
 DE Peptide #5040 encoded by breast cell single exon nucleic acid probe.  
 XX  
 KW Human; microarray; single exon probe; gene expression; breast;  
 KW disease; cancer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157271-A2.  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US000662.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 DR WPI; 2001-496933/54.  
 XX  
 PT New spatially-addressable set of single exon nucleic acid probes,  
 PT useful for measuring gene expression in sample derived from human  
 PT breast, comprises number of single exon nucleic acid probes -  
 XX  
 PS Claim 27; SEQ ID NO 15357; 327bp + sequence listing; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human breast and BT 474 cells. The method involves contacting

CC the probes with a collection of detectably labelled nucleic acids  
CC derived from mRNA of human breast, and then measuring the label  
CC bound to each probe of the microarray. The probes are useful for  
CC verifying the expression of regions of genomic DNA predicted to  
CC encode proteins. They are useful for gene discovery, and for  
CC determining predisposition and/or prognosing breast disease. Gene  
CC expression analysis is useful for assessing the toxicity of chemical  
CC agents on cells. The microarray of this invention presents a far greater  
CC diversity of probes for measuring gene expression, with far less bias  
CC than expressed sequence tag microarrays. The method is suitable for  
CC rapid production of functional information from genomic sequence. The  
CC present sequence is a peptide encoded by a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 50 AA;  
Query Match 35.6%; Score 262; DB 22; Length 50;  
Best Local Similarity 100.0%; Pred. No. 4.9e-22;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 26 RKKTFSLVHEVMAVENYAKDSLQWITDQYNKESDDKXHPRIFFVLKVRQ 75  
Db 1 RKKTFSLVHEVMAVENYAKDSLQWITDQYNKESDDKXHPRIFFVLKVRQ 50  
RESULT 12  
ABBI9524  
ID ABB19524 standard; Peptide; 50 AA.  
XX  
AC ABB19524;  
XX  
DT 04-FEB-2002 (first entry)  
XX  
DE Peptide #1592 encoded by human foetal liver single exon probe.  
XX  
KW Human; foetal liver; gene expression; single exon nucleic acid probe.  
XX  
OS Homo sapiens.  
XX  
PN WO200157277-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001MO-US00669.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
XX  
PR 26-MAY-2000; 2000US-0207456.  
XX  
PR 30-JUN-2000; 2000US-0608408.  
XX  
PR 03-AUG-2000; 2000US-0632366.  
XX  
PR 21-SEP-2000; 2000US-0234687.  
XX  
PR 27-SRP-2000; 2000US-0236359.  
XX  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLB-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-483447/52.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human foetal liver -  
XX  
PS Claim 27; SEQ ID NO 26721; 639pp + sequence listing; English.  
XX  
XX The invention relates to a single exon nucleic acid probe for  
XX measuring human gene expression in a sample derived from human foetal  
XX liver. The single exon nucleic acid probes may be used for predicting,  
XX measuring and displaying gene expression in samples derived from human  
XX foetal liver. The present sequence is a peptide encoded by a single exon  
XX nucleic acid probe of the invention.

CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 50 AA;  
Query Match 35.6%; Score 262; DB 22; Length 50;  
Best Local Similarity 100.0%; Pred. No. 4.9e-22;  
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 26 RKKTFSLVHEVMAVENYAKDSLQWITDQYNKESDDKXHPRIFFVLKVRQ 75  
Db 1 RKKTFSLVHEVMAVENYAKDSLQWITDQYNKESDDKXHPRIFFVLKVRQ 50  
RESULT 13  
ABBI9524  
ID ABB19524 standard; Protein; 50 AA.  
XX  
AC ABB19524;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Protein #1523 encoded by probe for measuring heart cell gene expression.  
XX  
KW Human; gene expression; heart; microarray; vascular system;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200157274-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001MO-US00666.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
XX  
PR 26-MAY-2000; 2000US-0207456.  
XX  
PR 30-JUN-2000; 2000US-0608408.  
XX  
PR 03-AUG-2000; 2000US-0632366.  
XX  
PR 21-SEP-2000; 2000US-0234687.  
XX  
PR 27-SRP-2000; 2000US-0236359.  
XX  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLB-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-48899/53.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
XX hearts -  
XX  
PS Claim 15; SEQ ID NO 21294; 530pp; English.  
XX  
XX The present invention relates to single exon nucleic acid probes for  
XX measuring human gene expression in a sample derived from human heart (see  
XX ABA21535-ABA41305). The present sequence is a protein encoded by one such  
XX probe. The probes may be used for predicting, measuring and displaying  
XX gene expression in samples derived from the human heart via microarrays.  
XX By measuring gene expression, the probes are useful for predicting,  
XX diagnosing, grading, staging, monitoring and prognosing diseases of the  
XX human heart and vascular system e.g. cardiovascular disease,  
XX hypertension, cardiac arrhythmias and congenital heart disease.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 50 AA;  
Query Match 35.6%; Score 262; DB 22; Length 50;  
Best Local Similarity 100.0%; Pred. No. 4.9e-22;

Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 RKKTFSLVHEVMAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKVOQ 75

Db 1 RKKTFSLVHEVMAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKVOQ 50

## RESULT 14

AA54871  
ID AA54871 standard; Protein; 50 AA.

AC AA54871;

DT 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe encoded protein SEQ ID NO: 26976.

XX Human; brain expressed exon; gene expression analysis; probe;

KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

XX eplepsy; cancer.

OS Homo sapiens.

PN MO200157275-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00667.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483446/52.

XX WPI; 2001-483446/52.

XX WPI; 2001-483446/52.

XX WPI; 2001-483446/52.

XX WPI; 2001-483446/52.

XX WPI; 2001-483446/52.

XX WPI; 2001-483446/52.

XX WPI; 2001-483446/52.

XX WPI; 2001-483446/52.

XX WPI; 2001-483446/52.

XX WPI; 2001-483446/52.

XX WPI; 2001-483446/52.

XX WPI; 2001-483446/52.

DT 06-NOV-2001 (first entry)

DE Human bone marrow expressed probe encoded protein SEQ ID NO: 27558.

XX Human; bone marrow expressed exon; gene expression analysis; probe;

KW microarray; cancer; leukaemia; lymphoma; myeloma.

XX Homo sapiens.

PN MO200157276-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00668.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

XX WPI; 2001-488900/53.

QY 26 RKKTFSLVHEVMAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKVOQ 75

Db 1 RKKTFSLVHEVMAVENYAKDSLQWITDQYNKESDDKXHFRIFRVLKVOQ 50

Search completed: January 21, 2004, 12:06:47  
Job time : 66.4064 secs



;; CURRENT FILING DATE: 2000-07-17  
;; PRIOR APPLICATION NUMBER: 09/431,480  
;; PRIOR FILING DATE: 1999-11-01  
;; PRIOR APPLICATION NUMBER: 60/109,217  
;; PRIOR FILING DATE: 1998-11-20  
;; PRIOR APPLICATION NUMBER: 60/156,382  
;; PRIOR FILING DATE: 1999-09-28  
;; NUMBER OF SEQ ID NOS: 22  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 3  
;; LENGTH: 142  
;; TYPE: PRT  
;; ORGANISM: Mus musculus  
US-09-617-302-3

Query Match 30.8%; Score 226.5; DB 3; Length 142;  
Best Local Similarity 35.1%; Pred. No. 1.6e-19;  
Matches 52; Conservative 34; Mismatches 39; Indels 23; Gaps 7;

QY 1 MAEP-WQALQLLAILTLTALPYQARKKTFLSVHEWAVENY-----AKDSLQW 49  
DB 1 MAKPLM--LSILITPIPALAVGDQSK-----NEVKA-QNYGGINISNANKQCVWF 51  
QY 50 ITDQYNKESDDKXHFRIFRVLKQVQVTDHLEHYLNVEMQWTTQCKP--ETTNCVPOER- 106  
DB 52 AMKEYNKESDDKXVFLVDVKILHAKLQITDRMEYQIDVQISRSNCKKPLMNTENCIPQKKP 111  
QY 107 ELHKQVNCFSVPVAPWPEQYKILNKSC 134  
DB 112 ELERKMCSFLVGLPWNGEFVMEKKC 139

RESULT 3  
US-09-431-480-4  
;; Sequence 4, Application US/09431480  
;; Patent No. 6235708  
;; GENERAL INFORMATION:  
;; APPLICANT: Holloway, James L.  
;; APPLICANT: Feldhaus, Andrew  
;; TITLE OF INVENTION: TESTIS SPECIFIC CYSTATIN-LIKE PROTEIN CYSTATIN T  
;; FILE REFERENCE: 98-72  
;; CURRENT APPLICATION NUMBER: US/09/431,480  
;; CURRENT FILING DATE: 1999-11-01  
;; EARLIER APPLICATION NUMBER: 60/109,217  
;; EARLIER FILING DATE: 1998-11-20  
;; EARLIER APPLICATION NUMBER: 60/156,382  
;; EARLIER FILING DATE: 1999-09-28  
;; NUMBER OF SEQ ID NOS: 22  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 4  
;; LENGTH: 142  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-431-480-4

Query Match 27.4%; Score 202; DB 3; Length 142;  
Best Local Similarity 36.3%; Pred. No. 1.4e-16;  
Matches 49; Conservative 29; Mismatches 51; Indels 6; Gaps 4;

QY 5 WQALQLLAILTLTALPYQARKKT--FLSVHEWAVENYAKDSLQWITDQYNKESDDKY 62  
DB 6 WLST--ILITPIPALVARKDPKNETGVRLKLPVNASNANKQCVLWPMQBYNKESDDKY 64  
QY 63 HFRIFRLKQVQVTDHLEHYLNVEMQWTTQCKP--ETTNCVPOER-ELHKQVNCFSVP 119  
DB 65 VFLVVKTLQAOQVTDHLEHYLNVEMQWTTQCKP--ETTNCVPOER-ELHKQVNCFSVP 124  
QY 120 AVPMFEOYKILNKSC 134  
DB 125 ALPMNGEFTVMEKKC 139

RESULT 4

US-09-617-302-4  
;; Sequence 4, Application US/09617302  
;; Patent No. 6245529  
;; GENERAL INFORMATION:  
;; APPLICANT: Holloway, James L.  
;; APPLICANT: Feldhaus, Andrew  
;; TITLE OF INVENTION: TESTIS SPECIFIC CYSTATIN-LIKE PROTEIN CYSTATIN T  
;; FILE REFERENCE: 98-72 C1  
;; CURRENT APPLICATION NUMBER: US/09/617,302  
;; CURRENT FILING DATE: 2000-07-17  
;; PRIOR APPLICATION NUMBER: 09/431,480  
;; PRIOR FILING DATE: 1999-11-01  
;; PRIOR APPLICATION NUMBER: 60/109,217  
;; PRIOR FILING DATE: 1998-11-20  
;; PRIOR APPLICATION NUMBER: 60/156,382  
;; PRIOR FILING DATE: 1999-09-28  
;; NUMBER OF SEQ ID NOS: 22  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 4  
;; LENGTH: 142  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-617-302-4

Query Match 27.4%; Score 202; DB 3; Length 142;  
Best Local Similarity 36.3%; Pred. No. 1.4e-16;  
Matches 49; Conservative 29; Mismatches 51; Indels 6; Gaps 4;

QY 5 WQALQLLAILTLTALPYQARKKT--FLSVHEWAVENYAKDSLQWITDQYNKESDDKY 62  
DB 6 WLST--ILITPIPALVARKDPKNETGVRLKLPVNASNANKQCVLWPMQBYNKESDDKY 64  
QY 63 HFRIFRLKQVQVTDHLEHYLNVEMQWTTQCKP--ETTNCVPOER-ELHKQVNCFSVP 119  
DB 65 VFLVVKTLQAOQVTDHLEHYLNVEMQWTTQCKP--ETTNCVPOER-ELHKQVNCFSVP 124  
QY 120 AVPMFEOYKILNKSC 134  
DB 125 ALPMNGEFTVMEKKC 139

RESULT 5  
US-09-431-480-2  
;; Sequence 2, Application US/09431480  
;; Patent No. 6235708  
;; GENERAL INFORMATION:  
;; APPLICANT: Holloway, James L.  
;; APPLICANT: Feldhaus, Andrew  
;; TITLE OF INVENTION: TESTIS SPECIFIC CYSTATIN-LIKE PROTEIN CYSTATIN T  
;; FILE REFERENCE: 98-72  
;; CURRENT APPLICATION NUMBER: US/09/431,480  
;; CURRENT FILING DATE: 1999-11-01  
;; EARLIER APPLICATION NUMBER: 60/109,217  
;; EARLIER FILING DATE: 1998-11-20  
;; EARLIER APPLICATION NUMBER: 60/156,382  
;; EARLIER FILING DATE: 1999-09-28  
;; NUMBER OF SEQ ID NOS: 22  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 2  
;; LENGTH: 141  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-09-431-480-2

Query Match 26.6%; Score 195.5; DB 3; Length 141;  
Best Local Similarity 31.2%; Pred. No. 8.2e-16;  
Matches 45; Conservative 30; Mismatches 42; Indels 27; Gaps 4;

QY 8 LQALLAILTL-----MALPYQARKKTFLSVHEWAVENYAKDSLQWITDQ 53  
DB 5 LQTLFLVITVEFVSRRVEMGSPQIVRPEDIPKSYV-----YVQHALWYAMKE 54  
QY 54 YNKESDDKXHFRIFRVLKQVQVTDHLEHYLNVEMQWTTQCK--PETTNCV-POERELHK 110

Db 55 YNPSNDLNFVRVVDILKSOEQTIDSLTYLEVNIARTMCKIAGDNENCLFOODPRMKK 114  
 QY 111 QVNCFPSVPAVPMFEOYKILNKSC 134  
 Db 115 MVFCIFVSSKPMKPELKMALKKQC 138

## RESULT 6

US-09-617-302-2  
 ; Sequence 2, Application US/09617302  
 ; Patent No. 6245529  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Holmway, James L.  
 ; APPLICANT: Feldhaus, Andrew  
 ; TITLE OF INVENTION: TESTIS SPECIFIC CYSTATIN-LIKE PROTEIN CYSTATIN T  
 ; FILE REFERENCE: 98-72 C1  
 ; CURRENT APPLICATION NUMBER: US/09/617,302  
 ; CURRENT FILING DATE: 2000-07-17  
 ; PRIOR APPLICATION NUMBER: 09/431,480  
 ; PRIOR FILING DATE: 1999-11-01  
 ; PRIOR APPLICATION NUMBER: 60/109,217  
 ; PRIOR FILING DATE: 1998-11-20  
 ; PRIOR APPLICATION NUMBER: 60/156,382  
 ; PRIOR FILING DATE: 1999-09-28  
 ; NUMBER OF SEQ ID NOS: 22  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 141  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-617-302-2

Query Match 26.6%; Score 195.5; DB 3; Length 141;  
 Best Local Similarity 31.2%; Pred. No. 8.2e-16;  
 Matches 45; Conservative 30; Mismatches 42; Indels 27; Gaps 4;

QY 8 LQLLAILLT-----MALPYQARKKTFLSVHEVNAVENYAKDSLQWITDQ 53  
 Db 5 LQTLFLVITVEFVSRRVEMSGPQIVRPEDIPKSYV-----YQHALMYAMKE 54  
 QY 54 YNPSNDLNFVRVVDILKSOEQTIDSLTYLEVNIARTMCKIAGDNENCLFOODPRMKK 110  
 Db 55 YNPSNDLNFVRVVDILKSOEQTIDSLTYLEVNIARTMCKIAGDNENCLFOODPRMKK 114  
 QY 111 QVNCFPSVPAVPMFEOYKILNKSC 134  
 Db 115 MVFCIFVSSKPMKPELKMALKKQC 138

## RESULT 7

US-09-886-319A-46  
 ; Sequence 46, Application US/09886319A  
 ; Patent No. 6586185  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wolf, Eckard  
 ; APPLICANT: Werner, Sabine  
 ; APPLICANT: Halle, Jörn-Peter  
 ; APPLICANT: Regenbogen, Johannes  
 ; APPLICANT: Goppelt, Andreas  
 ; TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for  
 ; TITLE OF INVENTION: the Diagnosis or Treatment of Skin Disorders and Wound  
 ; TITLE OF INVENTION: Healing and for the Identification of Pharmacologically  
 ; TITLE OF INVENTION: Active Substances  
 ; FILE REFERENCE: 50125/014002  
 ; CURRENT APPLICATION NUMBER: US/09/886,319A  
 ; CURRENT FILING DATE: 2001-06-20  
 ; PRIOR APPLICATION NUMBER: US 60/222,081  
 ; PRIOR FILING DATE: 2000-08-01  
 ; PRIOR APPLICATION NUMBER: DE 10030149.5  
 ; PRIOR FILING DATE: 2000-06-20  
 ; NUMBER OF SEQ ID NOS: 84  
 ; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 46  
 ; LENGTH: 140  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 ; US-09-886-319A-46

Query Match 25.3%; Score 186.5; DB 4; Length 140;  
 Best Local Similarity 30.9%; Pred. No. 9.7e-15;  
 Matches 43; Conservative 30; Mismatches 63; Indels 3; Gaps 2;

QY 1 MAEPQALDQLAILLTLMALPYQARKKTFLSVHEVNAVENYAKDSLQWITDQYNKESDD 60  
 Db 1 MASPLASLFLAVLAVANAAATPKQGPRLGABEADANBEGVRALDFAVSEYNGSND 60  
 QY 61 KYHFRIFRYLKVQROYTDHLHYLANVEMQWTTQCKET--TNC-VFOERELHQVNCFFS 117  
 Db 61 AYHSRAIOVVRARQOLVAGVNYFLDVEMGRITCTKSQTNLTDPCPHDQPHLMKALCSFQ 120  
 QY 118 VFAVPMFEOYKILNKSCSS 136  
 Db 121 IYSPWKGTHSLTKFSCKN 139

## RESULT 8

US-09-886-319A-48  
 ; Sequence 48, Application US/09886319A  
 ; Patent No. 6586185  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wolf, Eckard  
 ; APPLICANT: Werner, Sabine  
 ; APPLICANT: Halle, Jörn-Peter  
 ; APPLICANT: Regenbogen, Johannes  
 ; APPLICANT: Goppelt, Andreas  
 ; TITLE OF INVENTION: Use of Polypeptides or Nucleic Acids for  
 ; TITLE OF INVENTION: the Diagnosis or Treatment of Skin Disorders and Wound  
 ; TITLE OF INVENTION: Healing and for the Identification of Pharmacologically  
 ; TITLE OF INVENTION: Active Substances  
 ; FILE REFERENCE: 50125/014002  
 ; CURRENT APPLICATION NUMBER: US/09/886,319A  
 ; CURRENT FILING DATE: 2001-06-20  
 ; PRIOR APPLICATION NUMBER: US 60/222,081  
 ; PRIOR FILING DATE: 2000-08-01  
 ; PRIOR APPLICATION NUMBER: DE 10030149.5  
 ; PRIOR FILING DATE: 2000-06-20  
 ; NUMBER OF SEQ ID NOS: 84  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 48  
 ; LENGTH: 140  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 ; US-09-886-319A-48

Query Match 24.8%; Score 182.5; DB 4; Length 140;  
 Best Local Similarity 30.2%; Pred. No. 2.9e-14;  
 Matches 42; Conservative 31; Mismatches 63; Indels 3; Gaps 2;

QY 1 MAEPQALDQLAILLTLMALPYQARKKTFLSVHEVNAVENYAKDSLQWITDQYNKESDD 60  
 Db 1 MASPLASLFLAVLAVANAAATPKQGPRLGABEADANBEGVRALDFAVSEYNGSND 60  
 QY 61 KYHFRIFRYLKVQROYTDHLHYLANVEMQWTTQCKET--TNC-VFOERELHQVNCFFS 117  
 Db 61 AYHSRAIOVVRARQOLVAGVNYFLDVEMGRITCTKSQTNLTDPCPHDQPHLMKALCSFQ 120  
 QY 118 VFAVPMFEOYKILNKSCSS 136  
 Db 121 IYSPWKGTHSLTKFSCKN 139

## RESULT 9

US-09-431-480-5  
 ; Sequence 5, Application US/09431480  
 ; Patent No. 6235708

GENERAL INFORMATION:  
APPLICANT: Holloway, James L.  
TITLE OF INVENTION: TESTIS SPECIFIC CYSTATIN-LIKE PROTEIN CYSTATIN T  
FILE REFERENCE: 98-72  
CURRENT APPLICATION NUMBER: US/09/431,480  
CURRENT FILING DATE: 1999-11-01  
EARLIER APPLICATION NUMBER: 60/109,217  
EARLIER FILING DATE: 1998-11-20  
EARLIER APPLICATION NUMBER: 60/156,382  
EARLIER FILING DATE: 1999-09-28  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 5  
LENGTH: 140  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-431-480-5

Query Match 24.4%; Score 179.5; DB 3; Length 140;  
Best Local Similarity 30.2%; Pred. No. 6.7e-14;  
Matches 42; Conservative 30; Mismatches 64; Indels 3; Gaps 2;

QY 1 MAFBQALQLLALITLTMALPYQARKTFLSVHEVNAVENYAKDSIQWITDQYNKSSDD 60  
DB 1 MASPLRSLFLFLAVLGVMATPKQGRMLGAPEADANEGVRALDFAVSEYKNSND 60  
QY 61 KTHFRIRVLKVGQVTDHLEHYNEMQWTTCKPPT--TNC-VPOEREIHKQVNCFFS 117  
DB 61 AYHSRAIQVVARARQIVAGVNYFPDVEGRITTCYSQTNLTDCEPHDQPHLMRALCSFQ 120  
QY 118 VFAVPMFEQYKIINKSCSS 136  
DB 121 TVSVPMKGTSLTKFSCKN 139

RESULT 10  
US-09-617-302-5  
Sequence 5, Application US/09617302  
Patent No. 6245529  
GENERAL INFORMATION:  
APPLICANT: Holloway, James L.  
TITLE OF INVENTION: TESTIS SPECIFIC CYSTATIN-LIKE PROTEIN CYSTATIN T  
FILE REFERENCE: 98-72 CI  
CURRENT APPLICATION NUMBER: US/09/617,302  
CURRENT FILING DATE: 2000-07-17  
PRIOR APPLICATION NUMBER: 09/431,480  
PRIOR FILING DATE: 1999-11-01  
PRIOR APPLICATION NUMBER: 60/109,217  
PRIOR FILING DATE: 1998-11-20  
PRIOR APPLICATION NUMBER: 60/156,382  
PRIOR FILING DATE: 1999-09-28  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 5  
LENGTH: 140  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-617-302-5

Query Match 24.4%; Score 179.5; DB 3; Length 140;  
Best Local Similarity 30.2%; Pred. No. 6.7e-14;  
Matches 42; Conservative 30; Mismatches 64; Indels 3; Gaps 2;  
QY 1 MAFBQALQLLALITLTMALPYQARKTFLSVHEVNAVENYAKDSIQWITDQYNKSSDD 60  
DB 1 MASPLRSLFLFLAVLGVMATPKQGRMLGAPEADANEGVRALDFAVSEYKNSND 60  
QY 61 KTHFRIRVLKVGQVTDHLEHYNEMQWTTCKPPT--TNC-VPOEREIHKQVNCFFS 117  
DB 61 AYHSRAIQVVARARQIVAGVNYFPDVEGRITTCYSQTNLTDCEPHDQPHLMRALCSFQ 120

QY 118 VFAVPMFEQYKIINKSCSS 136  
DB 121 TVSVPMKGTSLTKFSCKN 139

RESULT 11  
5432264-6  
Patent No. 5432264  
APPLICANT: GRUBB, ANDERS; LUNDWALL, AKE; ABRAHAMSON, MAGNUS;  
TITLE OF INVENTION: RECOMBINANT 3-DES-OH-CYSTATIN C PRODUCED  
BY EXPRESSION IN A PROCARYOTIC HOST CELL  
NUMBER OF SEQUENCES: 8  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/929,290  
FILING DATE: 13-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 440,221  
FILING DATE: 21-NOV-1989  
APPLICATION NUMBER: 297,198  
FILING DATE: 20-MAY-1988  
SEQ ID NO: 6  
LENGTH: 146  
5432264-6

Query Match 24.2%; Score 178; DB 6; Length 146;  
Best Local Similarity 32.2%; Pred. No. 1.1e-13;  
Matches 47; Conservative 26; Mismatches 63; Indels 10; Gaps 4;

QY 1 MAFBQALQLLALITLTMALPYQARK----KTLFSVHEVNAVENYAKDSIQWITDQY 54  
DB 1 MAGPLRAPLILLLALVALAVSPAGSSPKGPRLVGPMDASVEEGVRALDFAVGEY 60  
QY 55 NKESDDKTHFRIRVLKVGQVTDHLEHYNEMQWTTCKP--PETNVC-VPOEREIHKQ 111  
DB 61 NKASNDMYHSRALQVVARARQIVAGVNYFPDVEGRITTCYQPNLDCPFHDQPHLMRK 120  
QY 112 VNCFFSVFAVPMFEQYKIINKSCSSD 137  
DB 121 AFGSFQYAVPM-QQMTLSKSTCQD 145

RESULT 12  
US-08-744-138-6  
Sequence 6, Application US/08744138  
Patent No. 6011012  
GENERAL INFORMATION:  
APPLICANT: Gentz, Reiner L.  
APPLICANT: Ni, Jian  
APPLICANT: Rosen, Craig A.  
APPLICANT: Yu, Guo-Liang  
TITLE OF INVENTION: Human Cystatin E  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/744,138  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PF202P1



LIBRARY: GenBank  
CLONE: 118195  
US-08-791-522-4

Query Match 23.6%; Score 173.5; DB 2; Length 139;  
Best Local Similarity 33.1%; Pred. No. 3.5e-13;  
Matches 43; Conservative 25; Mismatches 57; Indels 5; Gaps 3;

Qy 10 LLLAILTLMLPQARKKTFLSVHEVMAVEN--YAKDSLQMTDQYKESDDKXHPRI 67  
Db 9 VLLAAALMLVGAVIGSDESRRLGAPVPVDENDEGLQALQFAMAEYNRASNDKYSRRV 68

Qy 68 RVLKVRQVTDHLEHNLNVMQWTTQCKP--ETTNC-VPOREHLKQVNCFFSVFVAPWF 124  
Db 69 RVISAKQLVSGIKYILQVEIGRTTCPSGSDLOSCEFHDEPMAXTTCTFVVYSIPWL 128

Qy 125 EGYKILKSC 134  
Db 129 NQIKLESKC 138

RESULT 15

US-09-314-777-4  
Sequence 4, Application US/09314777  
Patent No. 6110686

GENERAL INFORMATION:

APPLICANT: Bandman, Olga

APPLICANT: Goll, Surya K.

TITLE OF INVENTION: NOVEL HUMAN CYSTATIN-LIKE

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESS: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/314,777

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/791,522

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0193 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 139 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: 118195

US-09-314-777-4

Query Match 23.6%; Score 173.5; DB 3; Length 139;  
Best Local Similarity 33.1%; Pred. No. 3.5e-13;  
Matches 43; Conservative 25; Mismatches 57; Indels 5; Gaps 3;

Qy 10 LLLAILTLMLPQARKKTFLSVHEVMAVEN--YAKDSLQMTDQYKESDDKXHPRI 67

Db 9 VLLAAALMLVGAVIGSDESRRLGAPVPVDENDEGLQALQFAMAEYNRASNDKYSRRV 68

Qy 68 RVLKVRQVTDHLEHNLNVMQWTTQCKP--ETTNC-VPOREHLKQVNCFFSVFVAPWF 124  
Db 69 RVISAKQLVSGIKYILQVEIGRTTCPSGSDLOSCEFHDEPMAXTTCTFVVYSIPWL 128

Qy 125 EGYKILKSC 134  
Db 129 NQIKLESKC 138

Search completed: January 21, 2004, 12:09:57  
Job time : 25.6323 secs



Pt		to inhibition of thrombotic events associated with cancer
Xx		
Pt	Claim 2; Page 96; 100pp; English.	
Xx		
Pt	The invention describes an isolated mammalian cystatin-8 (Zcys8)	
Xx	polypeptide (I). (I) is useful for: inhibiting cancer procoagulant	
Pt	protein in an individual and thus inhibiting the thrombotic events	
Xx	associated with cancer; promoting spermatogenesis; modulating seminal	
Pt	fluid viscosity; enhancing viability of cryopreserved sperm; sperm	
Xx	motility and fertilisation; and as antigenic peptides to generate	
Pt	antibodies. Zcys8 is useful as research reagent for characterising sites	
Xx	of interaction between Zcys8 and its receptor. Zcys8 is useful in	
Pt	enhancing fertilisation during assisted reproduction in humans and in	
Xx	animals. Anti-(I) antibodies are useful to screen biological samples like	
Pt	blood, urine, saliva, tissue biopsy and autopsy material in vitro for the	
Xx	presence of Zcys8. The antibodies are also useful to isolate large	
Pt	quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.	
Xx	The polynucleotide encoding (I) is useful to detect and to localise the	
Pt	expression of a Zcys8 gene in a biological sample and Zcys8	
Xx	oligonucleotide probes are useful for in vivo diagnosis. The	
Pt	polynucleotide encoding (I) is useful in determining whether a subject's	
Xx	chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene	
Pt	copy number changes, insertions, deletions, restriction site changes and	
Xx	rearrangements and genetic alterations that inactivate the Zcys8 gene.	
Pt	This sequence represents an antigenic peptide of human cystatin-8	
Xx	(Zcys8).	
Sq	Sequence 18 AA;	
Oy	Query Match 100.0%; Score 99; DB 23; Length 18;	
Dd	Best Local Similarity 100.0%; Pred. No. 1.9e-08;	
	Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0.	
	1 KDSLQWITDQYNKESDDK 18	
	1 KDSLQWITDQYNKESDDK 18	
RESULT 2		
ID AAU79858	AAU79858 standard; peptide; 35 AA.	
XX AC	AAU79858;	
XX DT	15-JUL-2002 (first entry)	
XX DE	Human cystatin-8 (Zcys8) antigenic fragment #6.	
XX KW	Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;	
XX RW	spermatogenesis; seminal fluid viscosity; cryopreserved sperm;	
XX OS	sperm motility; fertilisation; antigenic peptide.	
XX OS	Homo sapiens.	
XX PN	WO200220567-A2.	
XX PD	14-MAR-2002.	
XX PF	29-AUG-2001; 2001MO-US26868.	
XX PR	01-SEP-2000; 2000US-230230P.	
XX PA	(ZYMO ) ZYMOGENETICS INC.	
XX PI	Holloway JL, Gao Z, Bishop PD;	
XX DR	WPI; 2002-383044/41.	
PT Novel isolated mammalian cystatin-8 polypeptide useful for promoting		
PT spermatogenesis, and inhibiting cancer procoagulant protein which leads		
PT to inhibition of thrombotic events associated with cancer		
PS Claim 2; Page 96; 100pp; English.		

XX	The invention describes an isolated mammalian cystatin-8 (Zcys8)
CC	polypeptide (I). (I) is useful for: inhibiting cancer procoagulant
CC	protein in an individual and thus inhibiting the thrombotic events
CC	associated with cancer; promoting spermatogenesis, modulating seminal
CC	fluid viscosity, enhancing viability of cryopreserved sperm, sperm
CC	motility and fertilisation, and as antigenic peptides to generate
CC	antibodies. Zcys8 is useful as research reagent for characterising sites
CC	of interaction between Zcys8 and its receptor. Zcys8 is useful in
CC	enhancing fertilisation during assisted reproduction in humans and in
CC	animals. Anti-(I) antibodies are useful to screen biological samples like
CC	blood, urine, saliva, tissue biopsy and autopsy material in vitro for the
CC	presence of Zcys8. The antibodies are also useful to isolate large
CC	quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.
CC	The polynucleotide encoding (I) is useful to detect and to localise the
CC	expression of a Zcys8 gene in a biological sample and Zcys8
CC	oligonucleotide probes are useful for in vivo diagnosis. The
CC	polynucleotide encoding (I) is useful in determining whether a subject's
CC	chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene
CC	copy number changes, insertions, deletions, restriction site changes and
CC	rearrangements and genetic alterations that inactivate the Zcys8 gene.
CC	This sequence represents an antigenic peptide of human cystatin-8
CC	(Zcys8).
SQ	Sequence      35 AA;
	Query Match                  100.0%; Score 99; DB 23; Length 35;
	Best Local Similarity       100.0%; Pred. No. 4e-08;
Matches	18; Conservative     0; Mismatches     0; Indels     0; Gaps     0;
OY	1 KDSIQWITDOYNKESDDK 18          
Db	1 KDSLQWITDOYNKESDDK 18
	RESULT 3
ID	AAU79856
	AAU79856 standard; Peptide; 36 AA.
AC	AAU79856;
XX	15-JUL-2002 (first entry)
DE	Human cystatin-8 (Zcys8) antigenic fragment #4.
XX	Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;
KV	spermatogenesis; seminal fluid viscosity; cryopreserved sperm;
KM	sperm motility; fertilisation; antigenic fragment.
OS	Homo sapiens.
PN	WO200220567-A2.
PD	14-MAR-2002.
PF	29-AUG-2001; 2001WO-US26868.
PR	01-SEP-2000; 2000US-230230P.
PA	(ZYMO ) ZYMOGENETICS INC.
PI	Holloway JL, Gao Z, Bishop PD;
DR	WPI, 2002-383044/41.
PT	Novel isolated mammalian cystatin-8 polypeptide useful for promoting
PT	spermatogenesis, and inhibiting cancer procoagulant protein which leads
PS	to inhibition of thrombotic events associated with cancer -
XX	Claim 2; Page 95; 100p; English.
XX	The invention describes an isolated mammalian cystatin-8 (Zcys8)
CC	polypeptide (I). (I) is useful for: inhibiting cancer procoagulant

CC		protein in an individual and thus inhibiting the thrombotic events
CC		associated with cancer; promoting spermatogenesis, modulating seminal
CC		fluid viscosity, enhancing viability of cryopreserved sperm, sperm
CC		motility and fertilisation; and as antigenic peptides to generate
CC		antibodies. Zcy98 is useful as research reagent for characterising sites
CC		of interaction between Zcy98 and its receptor. Zcy98 is useful in
CC		enhancing fertilisation during assisted reproduction in humans and in
CC		animals. Anti-(I) antibodies are useful to screen biological samples like
CC		blood, urine, saliva, tissue biopsy and autopsy material in vitro for the
CC		presence of Zcy98. The antibodies are also useful to isolate large
CC		quantities of Zcy98 protein and DNA sequences that encode Zcy98 genes.
CC		The polynucleotide encoding (I) is useful to detect and to localise the
CC		expression of a Zcy98 gene in a biological sample and Zcy98
CC		oligonucleotide probes are useful for in vivo diagnosis. The
CC		polynucleotide encoding (I) is useful in determining whether a subject's
CC		chromosomes contain a mutation in the Zcy98 gene like aneuploidy, gene
CC		copy number changes, insertions, deletions, restriction site changes and
CC		rearrangements and genetic alterations that inactivate the Zcy98 gene.
CC		This sequence represents an antigenic fragment of human cystatin-B
CC		(Zcy98).
CC	SQ	Sequence      36 AA;
OY		Query Match                  100.0%; Score 99; DB 23; Length 36;
Dd		Best Local Similarity        100.0%; Pred. No. 4,1e-08;
		Matches    18; Conservative    0; Mismatches    0; Indels     0; Gaps    0;
		1 KDSLOWITDQYNKESDDK 18
		19 KDSLQWITDQYNKESDDK 36
RESULT 4		
ABG48915		ABG48915 standard; Peptide; 50 AA.
AC		ABG48915;
DT		25-FEB-2003 (first entry)
DE		Human liver peptide, SEQ ID NO 27563.
KW		Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KX		hypercholesterolaemia; coronary heart disease.
OS		Homo sapiens.
PN		WO200157273-A2.
PD		09-AUG-2001.
Pf		30-JAN-2001; 2001WO-US00664.
RX		04-FEB-2000; 2000US-0180312.
PR		26-MAY-2000; 2000US-0207456.
PR		30-JUN-2000; 2000US-0608408.
PR		03-AUG-2000; 2000US-0632366.
PR		21-SEP-2000; 2000US-0234687.
PR		27-SEP-2000; 2000US-0236359.
PR		04-OCT-2000; 2000GB-0024263.
PA		(MOLE-) MOLECULAR DYNAMICS INC.
PI		Penn SG, Hanzel DK, Chen W, Rank DR;
DR		WPI; 2001-488898/53.
PT		Human genome-derived single exon nucleic acid probes useful for
PT		analysing gene expression in human adult liver -
PS		Claim 27; SEQ ID NO 27563; 658pp; English.
CC		The invention relates to a single exon nucleic acid probe (SENP) (I) for

CC	measuring human gene expression in a sample derived from human adult
CC	liver, comprising one of 13109 defined nucleotide sequences given in the
CC	specification (or complements/ fragments). The probe hybridises at high
CC	stringency to a nucleic acid molecule expressed in the human adult
CC	liver. (I) may be used for predicting, measuring and displaying gene
CC	expression in samples derived from human adult liver. The genes
CC	identified may be involved in genetic liver diseases such as cirrhosis,
CC	hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolemia which
CC	is associated with coronary heart disease. ABG47348-ABG5930 represent
CC	human liver single exon encoded peptides of the invention.
CC	Note: The sequence information for this patent does not appear in the
CC	printed specification but was obtained in electronic format directly
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences.
SQ	Sequence 50 AA;
Query Match	100.0%; Score 99; DB 22; Length 50;
Best Local Similarity	100.0%; Pred. NO. 5.9e-09;
Matches 18; Conservative	0; Mismatches 0; Indels 0; Gaps 0
Oy	1 KDSLOWITDQYNKESDDK 18 
Db	19 KDSLQWITDQYNKESDDK 36
RESULT 5	
ABB28913	
ID	ABB28913 standard; Peptide; 50 AA.
AC	ABB28913;
XX	
DT	01-FEB-2002 (first entry)
XX	
DE	Peptide #1564 encoded by breast cell single exon nucleic acid probe.
XX	
KM	Human; microarray; single exon probe; gene expression; breast;
KW	disease; cancer.
XX	
OS	Homo sapiens.
FN	WO200157271-A2.
PM	
PD	09-AUG-2001.
XX	
PF	30-JAN-2001; 2001WO-US00662.
XX	
PR	04-FEB-2000; 2000US-0180312.
PR	26-MAY-2000; 2000US-0207456.
PR	30-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0632366.
PR	21-SEP-2000; 2000US-0234687.
PR	27-SEP-2000; 2000US-0236359.
PR	04-OCT-2000; 2000GB-0024263.
XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.
XX	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	
DR	WPI; 2001-496933/54.
XX	
PT	New spatially-addressable set of single exon nucleic acid probes,
PT	useful for measuring gene expression in sample derived from human
PT	breast, comprises number of single exon nucleic acid probes -
XX	
PS	Claim 27; SEQ ID NO 11881; 327pp + sequence listing; English.
XX	
CC	The invention relates to a spatially-addressable set of single exon
CC	nucleic acid probes for measuring gene expression in a sample derived
CC	from human breast and BT 474 cells. The method involves contacting
CC	the probes with a collection of detectably labelled nucleic acids
CC	derived from mRNA of human breast, and then measuring the label
CC	bound to each probe of the microarray. The probes are useful for
CC	verifying the expression of regions of genomic DNA predicted to

CC encode proteins. They are useful for gene discovery, and for  
CC determining predisposition and/or prognosing breast disease. Gene  
CC expression analysis is useful for assessing the toxicity of chemical  
CC agents on cells. The microarray of this invention presents a far greater  
CC diversity of probes for measuring gene expression, with far less bias  
CC than expressed sequence tag microarrays. The method is suitable for  
CC rapid production of functional information from genomic sequence. The  
CC present sequence is a peptide encoded by a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 50 AA;

Query Match 100.0%; Score 99; DB 22; Length 50;  
Best Local Similarity 100.0%; Pred. No. 5.9e-08;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KDSLQWITDQYNKESDDK 18  
Db 19 KDSLQWITDQYNKESDDK 36

RESULT 6  
ABR32389  
ID ABR32389 standard; Peptide; 50 AA.

XX AC ABR32389;  
XX DT 01-FEB-2002 (first entry)

DE Peptide #5040 encoded by breast cell single exon nucleic acid probe.

XX Human; microarray; single exon probe; gene expression; breast;  
XX disease; cancer.

XX OS Homo sapiens.

XX PN WO200157271-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00662.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-496933/54.

XX PT New spatially-addressable set of single exon nucleic acid probes,  
XX useful for measuring gene expression in sample derived from human  
XX breast; comprises number of single exon nucleic acid probes -

XX PS Claim 27; SEQ ID NO 15357; 327bp + sequence listing; English.

XX CC The invention relates to a spatially-addressable set of single exon  
XX nucleic acid probes for measuring gene expression in a sample derived  
XX from human breast and BR 474 cells. The method involves contacting  
XX the probes with a collection of detectably labelled nucleic acids  
XX derived from mRNA of human breast, and then measuring the label  
XX bound to each probe of the microarray. The probes are useful for  
XX verifying the expression of regions of genomic DNA predicted to  
XX encode proteins. They are useful for gene discovery, and for

CC determining predisposition and/or prognosing breast disease. Gene  
CC expression analysis is useful for assessing the toxicity of chemical  
CC agents on cells. The microarray of this invention presents a far greater  
CC diversity of probes for measuring gene expression, with far less bias  
CC than expressed sequence tag microarrays. The method is suitable for  
CC rapid production of functional information from genomic sequence. The  
CC present sequence is a peptide encoded by a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 50 AA;

Query Match 100.0%; Score 99; DB 22; Length 50;  
Best Local Similarity 100.0%; Pred. No. 5.9e-08;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KDSLQWITDQYNKESDDK 18  
Db 19 KDSLQWITDQYNKESDDK 36

RESULT 7  
ABR34086  
ID ABR34086 standard; Peptide; 50 AA.

XX AC ABR34086;  
XX DT 04-FEB-2002 (first entry)

DE Peptide #1592 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

XX OS Homo sapiens.

XX PN WO200157271-A2.

XX PD 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US00669.

XX PR 04-FEB-2000; 2000US-0180312.

XX PR 26-MAY-2000; 2000US-0207456.

XX PR 30-JUN-2000; 2000US-0608408.

XX PR 03-AUG-2000; 2000US-0632366.

XX PR 21-SEP-2000; 2000US-0234687.

XX PR 27-SEP-2000; 2000US-0236359.

XX PR 04-OCT-2000; 2000GB-0024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-483447/52.

XX PT Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human foetal liver -

XX PS Claim 27; SEQ ID NO 26721; 639bp + sequence listing; English.

XX CC The invention relates to a single exon nucleic acid probe for  
XX measuring human gene expression in a sample derived from human foetal  
XX liver. The single exon nucleic acid probes may be used for predicting,  
XX measuring and displaying gene expression in samples derived from human  
XX foetal liver. The present sequence is a peptide encoded by a single exon  
XX nucleic acid probe of the invention.  
XX Note: The sequence data for this patent did not form part of the  
XX printed specification, but was obtained in electronic format directly  
XX from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Query Match 100.0%; Score 99; DB 22; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-08;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KDSLOWITDOYNKESDDK 18  
 |||||  
 DB 19 KDSLOWITDOYNKESDDK 36

## RESULT 8

ABBI9524  
 ID ABBI9524 standard; Protein; 50 AA.

AC ABBI9524;

DT 23-JAN-2002 (first entry)

DE Protein #1523 encoded by probe for measuring heart cell gene expression.

KW Human; gene expression; heart; microarray; vascular system;  
 cardiovascular disease; hypertension; cardiac arrhythmia;

KM congenital heart disease.

XX Homo sapiens.

OS WO200157274-A2.

PN 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US00066.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;

PI WPI; 2001-488899/53.

DR Single exon nucleic acid probes for analyzing gene expression in human

XX hearts -

PS Claim 15; SEQ ID NO 21294; 530pp; English.

XX The present invention relates to single exon nucleic acid probes for

CC measuring human gene expression in a sample derived from human heart (see

CC ABBI9524-ABBI9524). The present sequence is a protein encoded by one such

CC probe. The probe may be used for predicting, measuring and displaying

CC gene expression in samples derived from the human heart via microarrays.

CC By measuring gene expression, the probes are useful for predicting,

CC diagnosing, grading, staging, monitoring and prognosing diseases of the

CC human heart and vascular system e.g. cardiovascular disease,

CC hypertension, cardiac arrhythmias and congenital heart disease.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WPI

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 50 AA;

XX Query Match 100.0%; Score 99; DB 22; Length 50;

XX Best Local Similarity 100.0%; Pred. No. 5.9e-08;

XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KDSLOWITDOYNKESDDK 18  
 |||||  
 DB 19 KDSLOWITDOYNKESDDK 36

RESULT 9  
 AAM54871  
 ID AAM54871 standard; Protein; 50 AA.

AC AAM54871;

DT 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe encoded protein SEQ ID NO: 26976.

KW Human; brain expressed exon; gene expression analysis; probe;

KM microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

KM epilepsy; cancer.

XX Homo sapiens.

OS WO200157275-A2.

PN 09-AUG-2001.

PD 30-JAN-2001; 2001WO-US00067.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;

PI WPI; 2001-483446/52.

DR Single exon nucleic acid probes for analyzing gene expression in human

XX brains -

PS Example 4; SEQ ID NO: 26976; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid

CC probes which are derived from genomic sequences expressed in the human

CC brain. They can be used to measure gene expression in brain cell samples,

CC which may enable the diagnosis and improved treatment of nervous system

CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,

CC epilepsy and cancers. The present sequence is a protein encoded by one of

XX the probes of the invention.

XX Sequence 50 AA;

XX Query Match 100.0%; Score 99; DB 22; Length 50;

XX Best Local Similarity 100.0%; Pred. No. 5.9e-08;

XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KDSLOWITDOYNKESDDK 18  
 |||||  
 DB 19 KDSLOWITDOYNKESDDK 36

RESULT 10  
 AAM67252  
 ID AAM67252 standard; Protein; 50 AA.

AC AAM67252;

DT 06-NOV-2001 (first entry)

DE Human bone marrow expressed probe encoded protein SEQ ID NO: 27558.

KW Human; bone marrow expressed exon; gene expression analysis; probe;

```

KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0633365.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
PS Example 4; SEQ ID NO: 27558; 658pp + Sequence listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention.
XX
SQ Sequence 50 AA;
Oy Query Match 100.0%; Score 99; DB 22; Length 50;
Best Local Similarity 100.0%; Pred. No. 5.9e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 KDSLQWITDQYKKSDDK 18
19 KDSLQWITDQYKKSDDK 36
RESULT 11
AAM15096
ID AAM15096 standard; Protein; 50 AA.
XX
AC AAM15096;
XX
DT 12-OCT-2001 (first entry)
DE Peptide #1530 encoded by probe for measuring cervical gene expression.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
cervical cancer.
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0633366.
PR 21-SEP-2000; 2000US-0234687.

```

```

PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
    analyzing gene expression in human cervical epithelial cells -
PS Claim 27; SEQ ID No 19922; 487bp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SNP: see AAI10068-AA128459). The present sequence is a peptide encoded
CC by one such probe. The SNPs are derived from human HeLa cells. The SNPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 50 AA;
Query Match 100.0%; Score 99; DB 22; Length 50;
Best Local Similarity 100.0%; Pred. No. 5,9e-08;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
OY 1 KDSIQWITDQYNKESDDX 18
   |||||
DB 19 KDSIQWITDQYNKESDDX 36
RESULT 12
AAM27545
ID AAM27545 standard; Protein; 50 AA.
XX AC AAM27545;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #1582 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for

```

PT analyzing gene expression in human placenta -  
 CC Claim 27; SEQ ID No 27814; 654pp; English.  
 XX  
 CC The present invention relates to single exon nucleic acid probes (SENP:  
 CC see A100010-A110067). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for producing a microarray for  
 CC predicting, measuring and displaying gene expression in samples derived  
 CC from human placenta. The probes are useful for antenatal diagnosis of  
 CC human genetic disorders.  
 XX  
 SQ Sequence 50 AA;  
 QY  
 Query Match 100.0%; Score 99; DB 22; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-08;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 1 KDSLQWITTDQYNKESDDK 18  
 19 KDSLQWITTDQYNKESDDK 36  
 RESULT 13  
 AAM02833  
 ID AAM02833 standard; Protein; 50 AA.  
 AC AAM02833;  
 XX  
 DT 09-OCT-2001 (first entry)  
 DE Peptide #1515 encoded by probe for measuring breast gene expression.  
 XX Probe; human; breast disease; breast cancer; development disorder;  
 KM inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157270-A2.  
 PD 09-AUG-2001.  
 XX  
 PF 29-JAN-2001; 2001WO-US00661.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2001-476286/51.  
 XX  
 PT Novel single exon nucleic acid probe used to measuring gene expression  
 PT in a human breast -  
 XX  
 PS Claim 27; SEQ ID No 11573; 322pp; English.  
 XX  
 CC The present invention relates to novel single exon nucleic acid probes  
 CC (see A100010-A110067). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for measuring human gene expression in  
 CC a human breast sample, where the probe hybridizes at high stringency to a  
 CC nucleic acid expressed in the human breast. The probes are useful for  
 CC predicting, diagnosing, grading, staging, monitoring and prognosing  
 CC diseases of the human breast, particularly those diseases with polygenic  
 CC aetiology. The diseases include: breast cancer; disorders of development,  
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative  
 CC breast disease and non-carcinoma tumours.  
 CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 50 AA;  
 QY  
 Query Match 100.0%; Score 99; DB 22; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-08;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 1 KDSLQWITTDQYNKESDDK 18  
 19 KDSLQWITTDQYNKESDDK 36  
 RESULT 14  
 ABG36903  
 ID ABG36903 standard; Peptide; 50 AA.  
 AC ABG36903;  
 XX  
 DT 19-AUG-2002 (first entry)  
 DE Human peptide encoded by genome-derived single exon probe SEQ ID 26568.  
 XX  
 KM Human; single exon probe; asthma; lung cancer; COPD; ILD;  
 KM chronic obstructive pulmonary disease; interstitial lung disease;  
 KM familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KM tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KM Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KM pulmonary histiocytosis; lymphangioleiomyomatosis; Karsagen syndrome;  
 KM pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KM primary ciliary dyskinesia; pulmonary hypertension;  
 KM hyaline membrane disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200186003-A2.  
 PD 15-NOV-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00665.  
 XX  
 PR 04-FEB-2000; 2000US-180312P.  
 PR 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2002-114183/15.  
 XX  
 PT Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples -  
 XX  
 PS Claim 27; SEQ ID No 26568; 634pp; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridize at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising

(a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarray having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene, a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemorrhoidosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karsenger syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a peptide/protein encoded by a single exon probe of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

CC Sequence 50 AA;

CC Query Match 100.0%; Score 99; DB 23; Length 50;

CC Best Local Similarity 100.0%; Pred. No. 5.9e-08;

CC Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDSLQWITDQYNKESDDK 18

DB 19 KDSLQWITDQYNKESDDK 36

RESULT 15

AAU79853 AAV79853 standard; Protein; 115 AA.

AC AAU79853;

DT 15-JUL-2002 (first entry)

DE Human cystatin-8 (Zcys8) antigenic fragment #1.

KM Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis; spermatogenesis; seminal fluid viscosity; cryopreserved sperm; sperm motility; fertilisation; antigenic fragment.

OS Homo sapiens.

PN WO200220567-A2.

PD 14-MAR-2002.

PF 29-AUG-2001; 2001WO-US26868.

PR 01-SEP-2000; 2000US-230230P.

PA (ZYMO ) ZYMOGENETICS INC.

PI Holloway JL, Gao Z, Bishop PD;

DR WPI; 2002-383044/41.

PT Novel isolated mammalian cystatin-8 polypeptide useful for promoting spermatogenesis, and inhibiting cancer procoagulant protein which leads to inhibition of thrombotic events associated with cancer

XX Claim 2; Page 94; 100pp; English.

PS The invention describes an isolated mammalian cystatin-8 (Zcys8) polypeptide (1). (1) is useful for: inhibiting cancer procoagulant protein in an individual and thus inhibiting the thrombotic events associated with cancer; promoting spermatogenesis; modulating seminal fluid viscosity, enhancing viability of cryopreserved sperm, sperm motility and fertilisation; and as antigenic peptides to generate antibodies. Zcys8 is useful as research reagent for characterising sites of interaction between Zcys8 and its receptor. Zcys8 is useful in enhancing fertilisation during assisted reproduction in humans and in animals. Anti-(1) antibodies are useful to screen biological samples like blood, urine, saliva, tissue biopsy and autopsy material in vitro for the presence of Zcys8. The antibodies are also useful to isolate large quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.

CC The polynucleotide encoding (1) is useful to detect and to localise the expression of a Zcys8 gene in a biological sample and Zcys8 oligonucleotide probes are useful for in vivo diagnosis. The polynucleotide encoding (1) is useful in determining whether a subject's chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene copy number changes, insertions, deletions, restriction site changes and rearrangements and genetic alterations that inactivate the Zcys8 gene.

CC This sequence represents an antigenic fragment of human cystatin-8 (Zcys8).

CC Sequence 115 AA;

CC Query Match 100.0%; Score 99; DB 23; Length 115;

CC Best Local Similarity 100.0%; Pred. No. 1.5e-07;

CC Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDSLQWITDQYNKESDDK 18

DB 22 KDSLQWITDQYNKESDDK 39

Search completed: January 21, 2004, 12:06:47  
Job time : 8.59355 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 12:00:59 ; Search time 2.32258 Seconds  
(without alignments)  
364.457 Million cell updates/sec

Title: US-09-941-314-7

Perfect score: 99  
Sequence: 1 KDSLQWITTDQYNKESDK 18

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	99	100.0	CS11_HUMAN	Q9H112 homo sapien
2	63.6	137	CS11_MOUSE	Q94269 mus musculu
3	47	139	CS11_HUMAN	O60676 homo sapien
4	47	142	NCAP_HUMAN	P04873 bunyavirus
5	47	235	MESJ_HABIN	P44689 haemophilus
6	46.5	430	SVR STRAM	Q99W75 staphylococ
7	45.5	807	HYSA_STAU	O59801 staphylococ
8	45	333	SVRA_TREPA	O99832 ureaplasma
9	44.4	60	Y02B_BPT4	P39232 bacterioph
10	44.4	168	BIP3_TOBAC	Q03683 nicotiana t
11	44.4	270	YTWK_BACSU	O34852 bacillus su
12	44.4	290	BIP1_TOBAC	Q03681 nicotiana t
13	44.4	292	BIP2_TOBAC	Q03682 nicotiana t
14	44.4	293	BIP8_TOBAC	Q03686 nicotiana t
15	44.4	403	PRE1_STAU	P03857 staphylococ
16	44.4	663	BIP2_MAIZE	P24067 zea mays (m
17	44.4	663	BIP3_MAIZE	O24581 zea mays (m
18	44.4	666	BIP1_ICYES	P49118 lycopersico
19	44.4	667	BIP4_TOBAC	Q03684 nicotiana t
20	44.4	668	BIP5_TOBAC	Q03685 nicotiana t
21	44.4	668	BIP1_STEOL	Q42434 spinacia ol
22	44.4	2258	FAS_PIG	O951D1 sus scrofa
23	43.4	142	CS18_MOUSE	P31276 mus musculu
24	43.4	190	RRE_FUSNN	O81529 fusobacteri
25	43.4	366	APM1_YEAST	P22936 saccharomyc
26	43.4	485	SVR_BACHD	O95966 bacillus ha
27	43.4	504	GUMW_ERMCA	O59395 erwinia car
28	43.4	783	YAYB_SCHPO	P10218 schizosach
29	42.4	116	CYT_COTJA	O81061 cucurbitac
30	42.4	240	SODE_HUMAN	P08294 homo sapien
31	42.4	264	TVSY_XANAC	O89P46 xanthomonas
32	42.4	264	TVSY_XANCP	O89C57 xanthomonas
33	42.4	442	TRME_MYCGE	P47254 mycoplasma

34	42	42.4	668	1	BIP2_ARATH	Q93043 arabidopsis
35	42	42.4	669	1	BIP1_ARATH	Q91K33 arabidopsis
36	42	42.4	976	1	SCP1_HUMAN	O15431 homo sapien
37	41	41.4	156	1	MIRC_PATYE	P05944 patinopete
38	41	41.4	156	1	MUR_CHLNT	P05963 chlamys nup
39	41	41.4	354	1	YCGF_HABIN	P43931 haemophilus
40	41	41.4	460	1	BE1A_YARLI	O59949 yarrowia 11
41	41	41.4	529	1	TIM_SALPO	P07989 salmoneilla
42	41	41.4	656	1	PEM1_CAREL	P17221 caenorhabdi
43	41	41.4	880	1	GUN4_THERU	P26221 thermomonas
44	41	41.4	1181	1	MYSC_DICDI	P42522 dictyosteli
45	40.5	40.9	524	1	YAB4_MYCPN	P75609 mycoplasma

## ALIGNMENTS

RESULT 1  
ID CS11\_HUMAN STANDARD; PRT; 137 AA.  
AC Q9H112; Q9H113;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Cystatin 11 precursor.  
GN CS11 OR CSTBL.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NCBI\_Taxid=9606;  
[1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RX MEDLINE=21636749; PubMed=11780052;  
RA DeLoukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggaley C.L.,  
Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
Buck D., Burrill W.D., Butler A.P., Carter C., Carter N.P.,  
Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
Clegg S., Cobley V.B., Collier R.E., Connor R.B., Corby N.R.,  
Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,  
Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
Layvaesialho M.H., Leversha M.A., Lloyd C., Lloyd G.K., Lovell J.D.,  
Marsh V.L., Martin S.L., McConachie L.J., McElay K., McMurtry A.A.,  
Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H.,  
Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhen R., Sims S.,  
Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sultson J.E.,  
Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,  
Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
Rogers J.;  
RT The DNA sequence and comparative analysis of human chromosome 20.;  
NA Nature 414:865-871(2001).  
CC -|- SUBCELLULAR LOCATION: Secreted (potential).  
CC -|- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=Q9H112-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q9H112-2; Sequence=VSP\_001260;  
CC Note=No experimental confirmation available;  
CC -|- SIMILARITY: BELONGS TO THE CYSTATIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; AL096677; CAC13170.1; -  
 DR EMBL; AL096677; CAC17423.1; -  
 DR HSSP; P01038; IA90.  
 DR HGNC; HGNC:15959; CST11.  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF00031; Cystatin; 1.  
 DR SMART; SM00043; CY; 1.  
 DR PROSITE; PS00287; CYSTATIN; FALSE NEG.  
 KW Thiol protease inhibitor; Signal; Alternative splicing.  
 FT SIGNAL 1 25  
 FT CHAIN 26 137  
 FT SITE 75 79  
 FT DISULFID 93 101  
 FT DISULFID 114 134  
 FT CARBOHYD 131 131  
 FT VARPELIC 76 110  
 SQ SEQUENCE 137 AA; 16375 MW; C585C8C39A585C3B CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 99; DB 1; Length 137;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDSLOWITDQYNKESDDK 18  
 DB 44 KDSLOWITDQYNKESDDK 61

RESULT 2  
 CS11\_MOUSE STANDARD; PRT; 139 AA.  
 AC Q9D269;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Cystatin 11 precursor.  
 GN CST11.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OK NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=EpIdidymis;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Araiawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishii K., Kiyosawa H., Kondo S., Yamamata I.,  
 RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Pleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,  
 RA Kuehl P., Lewis S., Matsumoto Y., Nishikido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Stabili F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Guetlinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons F., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,  
 RA Notone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyszynski A., Yoshida K., Hasegawa Y., Kawai H., Kohseki S.,  
 RA Hayashizaki Y.,  
 RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 403:685-690(2001).  
 CC -1- SUBCELLULAR LOCATION: Secreted (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE CYSTATIN FAMILY.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 DR EMBL; AK020300; BAB32061.1; -  
 DR HSSP; P01034; IG96.  
 DR WGI; WGI:1925490; Cct11.  
 DR InterPro; IPR000010; Cystatin.  
 DR Pfam; PF00031; Cystatin; 1.  
 DR SMART; SM00043; CY; 1.  
 DR PROSITE; PS00287; CYSTATIN; FALSE NEG.  
 KW Thiol protease inhibitor; Signal.  
 FT SIGNAL 1 28  
 FT CHAIN 29 139  
 FT SITE 76 80  
 FT DISULFID 94 102  
 FT DISULFID 115 135  
 FT CARBOHYD 134 134  
 SQ SEQUENCE 139 AA; 16217 MW; F228D9815FA32640 CRC64;

Query Match  
 Best Local Similarity 63.6%; Score 63; DB 1; Length 139;  
 Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDSLOWITDQYNKESDD 17  
 DB 45 KETLEYVTDEYNKESDD 61

RESULT 3  
 CST8\_HUMAN STANDARD; PRT; 142 AA.  
 AC O60676;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cystatin-related epididymal spermatogenic protein precursor (Cystatin  
 DE 8).  
 GN CST8 OR CRS.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OK NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=95344753; PubMed=7619504;  
 RA Cornwall G.A., Hann S.R.,  
 RA "Transient appearance of CRS protein during spermatogenesis and  
 RA caput epididymal sperm maturation."  
 RL Mol. Reprod. Dev. 41:37-46(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=21638749; PubMed=11780052;  
 RX Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrakas G., Almeida J.P., Babbage A.K., Baggaley C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
 RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burrill W.D., Butler A.P., Carter C., Carter N.P.,  
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.B., Collier R.E., Connor R.E., Corby N.R.,  
 RA Coulson A.G., Coville G.J., Deadman R., Dhami P.D., Dunn M.,  
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Grahnam D.V., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle B., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights K., Lawlor S., Lawlor S.,  
 RA Leveasialho M.H., Levesha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McConachie L.J., McElay K., McMurray A.A.,

```

RA MEDLINE=83164355; PubMed=6834480;
RX Akeshi H., Bishop D.H.L.;
RT "Comparison of the sequences and coding of La Crosse and snowshoe
RL hare bunyavirus S RNA species.";
RN J. Virol. 45:1155-1158(1983).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=83303821; PubMed=6684362;
RA Cabrera A.C.D. Jr., Holloway B.P., Obijeski J.F.;
RT "Molecular cloning and sequencing of the La Crosse virus S RNA.";
RL Virology 128:463-468(1983).
CC -1 SUBCELLULAR LOCATION: INTERNAL PROTEIN OF VIRUS PARTICLE.
CC -1 SIMILARITY: 45% IDENTITY TO AINO VIRUS NUCLEOPROTEIN N AND TO
CC SNOWSHOE HARE VIRUS NUCLEOPROTEIN N.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on ways
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1db-sib.ch/announce/).
CC or send an email to license@1db-sib.ch.
CC -----
CC EMBL: K00610; AAA42782.1; -.
CC DR EMBL: K00108; AAA43779.1; -.
CC DR PIR: A04104; VHVULV.
CC DR InterPro: IPR001784; Bunya_nucleocap.
CC DR Pfam: PF00952; Bunya_nucleocap; 1.
CC DR ProDom: PD001909; Bunya_nucleocap; 1.
CC KM Nucleocapsid; RNA-binding.
CC
CC SO SEQUENCE 235 AA; 26530 MW; 56EBB4D64AD0A96 CRC64;
CC
Qy 5 QWITDQYNKESDD 17
Db 101 RWVLDDQYNENDDE 113
CC
RESULT 5
MEST_HABIN STANDARD; PRT; 430 AA.
AC P44689;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative cell cycle protein mest homolog.
GN MEST OR H10404.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBT_TaxId=727;
OX [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Rd / KM20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness B.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McElaney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom B., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Pritchman J.L., Fuhrmann J.L., Georgagen N.S.M.,
RA Gehm C.L., McDonald L.A., Small K.V., Frazer C.M., Smith H.O.,
RA Venter J.C.;
RT Rd.;
RT Science 269:496-512(1995).
CC -1 SIMILARITY: BELONGS TO THE UPF0072 (MEST/VCF62) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL: U32723; AAC22063.1; -  
 CC PIR: D64151; D64151.  
 CC TIGR: H10404; -  
 CC InterPro: IPR000541; UPF0021.  
 CC Pfam: PF01171; ATP\_bind3; 1.  
 CC Complete proteome.  
 CC SEQUENCE 430 AA; 50052 MW; FPD0831DB67C1C6C CRC64;

QY Query Match 47.5%; Score 47; DB 1; Length 430;  
 Best Local Similarity 46.7%; Pred. No. 7.8;  
 Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

DB 1 KDSLOWITDOYNKES 15  
 175 KEKLMWITDESNEDN 189

RESULT 6  
 SVE\_STAM STANDARD; PRT; 484 AA.

AC Q99W75;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Glutamy1-CRNA synthetase (EC 6.1.1.17) (Glutamate--tRNA ligase)  
 DE (GLURS)  
 GN GLIX OR SAV0528 OR SA0486 OR MW0483.  
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699),  
 OS Staphylococcus aureus (strain N315), and  
 OS Staphylococcus aureus (strain MM2).  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 CX NCBI\_TaxID=158878; 158879, 196620;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Mu50 / ATCC 700699, and N315;  
 RX MEDLINE=21311952; PubMed=11418146;  
 RA Kuroda M., Ohba T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-Q., Ito T.,  
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
 RA Mizutani-U Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
 RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,  
 RA Kanehisa M., Yamashita A., Oshida K., Furuya K., Yoshino C., Shiba T.,  
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;  
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
 RT aureus.";  
 RL Lancet 357:1225-1240(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MM2;  
 RX MEDLINE=22040717; PubMed=12044378;  
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
 RA Nagai Y., Iwana N., Asano K., Naimi T., Kuroda H., Cui L.,  
 RA Yamamoto K., Hiramatsu K.;  
 RT "Genome and virulence determinants of high virulence community-  
 RT acquired MRSA.";  
 RL Lancet 355:1819-1827(2002).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +  
 CC dihydrophate + L-glutamyl-tRNA(Glu).  
 CC -1- SUBUNIT: Monomer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: Belongs to class-I aminocyl-tRNA synthetase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL: AP003359; BAB56690.1; -  
 CC EMBL: AP003130; BAB41716.1; -  
 CC EMBL: AP004823; BAB94348.1; -  
 CC PIR: A89820; A89820.  
 CC HSSP: P27000; 1GLN.  
 CC HAMAP: MF\_00022; -; 1.  
 CC InterPro: IPR004527; GLIX\_bact.  
 CC InterPro: IPR000924; Glu\_tRNA-synt\_1c.  
 CC InterPro: IPR001412; tRNA-synt\_1.  
 CC Pfam: PF00749; tRNA-synt\_1c; 1.  
 CC PRINTS: PR00987; TRNASYNTGLU.  
 CC TIGRFAMs: TIGR00464; gltx\_bact; 1.  
 CC PROSITE: PS00178; AA\_tRNA\_LIGASE\_1; 1.  
 CC Aminoacyl-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding;  
 CC Complete proteome.  
 CC SITE 11 "HIGH" REGION.  
 CC SITE 252 "KMSKS" REGION.  
 CC BINDING 255 ATP (BY SIMILARITY).  
 CC SEQUENCE 484 AA; 56288 MW; 4CBASFF08DA33EFA CRC64;

QY Query Match 46.5%; Score 46; DB 1; Length 484;  
 Best Local Similarity 43.8%; Pred. No. 13;  
 Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

DB 1 KDSLOWITDOYNKESD 16  
 318 KQKLAWVNNQYMKOD 333

RESULT 7  
 HISA STAU STANDARD; PRT; 807 AA.

AC Q59801;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Hyaluronate lyase precursor (EC 4.2.2.1) (Hyaluronidase) (HYase).  
 GN HISA.  
 OS Staphylococcus aureus.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 CX NCBI\_TaxID=1280;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCTC 8325-4;  
 RX MEDLINE=96009286; PubMed=7557301;  
 RA Farrell A.M., Taylor D., Holland K.T.;  
 RT "Cloning, nucleotide sequence determination and expression of the  
 RT Staphylococcus aureus hyaluronate lyase gene.";  
 RL FEMS Microbiol. Lett. 130:81-85(1995).  
 CC -1- CATALYTIC ACTIVITY: Hyaluronate = N 3-(4-deoxy-beta-D-gluc-4-  
 CC enuronosyl)-N-acetyl-D-glucosamine.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 8.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL: U21221; AAA82984.1; -  
 CC InterPro: IPR003159; Lyase\_8.  
 CC InterPro: IPR004103; Lyase\_8\_C.  
 CC Pfam: PF02278; Lyase\_8; 1.  
 CC Pfam: PF02884; Lyase\_8\_C; 1.  
 CC Lyase; Signal.  
 CC SIGNAL 1 40 POTENTIAL.

FT CHAIN 41 807 HYALURONATE LYASE.  
SQ SEQUENCE 807 AA; 91984 MW; FFE88BDC07418B84 CRC64;

Query Match 46.0%; Score 45.5; DB 1; Length 807;  
Best Local Similarity 47.4%; Pred. No. 26;  
Matches 9; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

Qy 1 KDSLOWITDQYNKESDDK 18  
Db 145 KDALEWIKHNAVYKXEPDKK 163

## RESULT 8

SYRA\_UREPA STANDARD; PRT; 333 AA.

AC 09PQ32;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Phenylalanyl-tRNA synthetase alpha chain (EC 6.1.1.20)  
DE (Phenylalanyl-tRNA ligase alpha chain) (PheRS).  
GN PHE8 OR U0458.

OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.

NCBI\_TaxID=134821;

SEQUENCE FROM N.A.

STRAIN=Serovar 3;  
MEDLINE=20500219; PubMed=11048724;

Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,

Caswell G.H.;  
"The complete sequence of the mucosal pathogen Ureaplasma

urealyticum.";  
Nature 407:757-762(2000).

-1 CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +  
diphosphate + L-phenylalanyl-tRNA(Phe).

-1 COFACTOR: Binds 2 magnesium ions per tetramer (By similarity).  
-1 SUBUNIT: Tetramer of two alpha and two beta chains (By

similarity).

-1 SUBCELLULAR LOCATION: Cytoplasmic.  
-1 SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.  
Phe-tRNA synthetase alpha chain subfamily 1.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AE002142; AAF30870.1; -

HSSP; P27001; 1PYS.

HAMAP; MF\_00281; -; 1.

InterPro; IPR004529; PheS.

InterPro; IPR002319; tRNA-synt\_2d.

InterPro; IPR006195; tRNA\_ligase\_II.

Pfam; PF01409; tRNA-synt\_2d; 1.

TIGRFAMs; TIGR00468; pheS; 1.

PROSITE; PS00862; AA-TRNA\_LIGASE\_II; 1.

Aminoacyl-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding;

Metal-binding; Magnesium; Complete proteome.

METAL 248  
SEQUENCE 333 AA; 38712 MW; SCAB3530FA34PFCF CRC64;

Query Match 45.5%; Score 45; DB 1; Length 333;  
Best Local Similarity 44.4%; Pred. No. 12;

Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 1 KDSLOWITDQYNKESDDK 18  
Db 59 KQGIETFTDQILKELNDK 76

## RESULT 9

Y02B\_BPT4 STANDARD; PRT; 60 AA.

AC P39232;

DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hypothetical 7.0 kDa protein in sp-Op41 intergenic region.  
GN Y02B OR 61.5.

OS Bacteriophage T4.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;  
OC T4-like viruses.

NCBI\_TaxID=10665;

SEQUENCE FROM N.A.

MEDLINE=93188183; PubMed=8383243;  
RA Selick H.E., Stormo G.D., Dyson R.L., Alberts B.M.;

"Analysis of five presumptive protein-coding sequences clustered  
RT between the primosome genes, 41 and 61, of bacteriophages T4, T2, and  
RT T6.";  
J. Virol. 67:2305-2316(1993).

NCBI\_TaxID=10665;

SEQUENCE FROM N.A.

Kutner E., Arisaka F., Kunisawa T., Tsugita A., Mosig G.,  
Mezianzhinov V., Ruger W., Stidham T., Thomas E.;

"Bacteriophage T4 genome analysis.";  
Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; S57514; AAB25708.1; -

EMBL; AF158101; AAD42512.1; -

DR PIR; A45681; A45681.  
KW Hypothetical protein.

SEQUENCE 60 AA; 7027 MW; A726546B9AC74B2 CRC64;

Query Match 44.4%; Score 44; DB 1; Length 60;  
Best Local Similarity 46.2%; Pred. No. 2.7;

Matches 12; Conservative 1; Mismatches 5; Indels 8; Gaps 2;

Qy 1 KDSLOWITDQYNKESDDK 18  
Db 26 KDVQWATTOYABVNAKALVKAFTDDK 51

## RESULT 10

BIP3\_TOBAC STANDARD; PRT; 168 AA.

AC Q03683;

DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Luminal binding protein 3 (Bip 3) (78 kDa glucose-regulated protein  
homolog 3) (GRP 78-3) (Fragment).

GN BIP3.

OS Nicotiana tabacum (Common tobacco).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; Lamiales; Solanales; Solanaceae; Nicotiana.

NCBI\_TaxID=4097;

SEQUENCE FROM N.A.

MEDLINE=92361242; PubMed=1822990;  
RA Denicke J., Goldman M.H., Demolder J., Seurinck J., Botterman J.;

"The tobacco luminal binding protein is encoded by a multigene  
RT family.";

```

[2] Plant Cell 3:1025-1035(1991).
RN RA ERRATUM.
RP Demecke J., Goldman M.H., Demolder J., Seurinck J., Botterman J.;
RA Plant Cell 3:1251-1251(1991).
CC -1- FUNCTION: PROBABLY PLAYS A ROLE IN FACILITATING THE ASSEMBLY OF
CC MULTIMERIC PROTEIN COMPLEXES INSIDE THE ER.
CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
-----
CC CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isdb-sib.ch/announce/
CC or send an email to license@isdb-sib.ch).
-----
DR DR EMBL; X60061; CAA442663.1; -.
DR HSBP, P04475; IDKY.
DR InterPro; IPR000886; ER_target.
DR InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; HSP70_1.
DR Prodom; PD000089; Hsp70_1.
DR PROSITE; PS00297; HSP70_1; PARTIAL.
DR PROSITE; PS00329; HSP70_2; PARTIAL.
DR PROSITE; PS01036; HSP70_3; PARTIAL.
DR PROSITE; PS00014; ER_TARGET; 1.
KW ATP-binding; Endoplasmic reticulum; Multigene family.
FT NON TER
FT CAROXYD 120 120 N-LINKED (GLCNAC... ) (POTENTIAL).
FT SITE 165 168 PREVENT SECRETION FROM ER (POTENTIAL).
SQ SEQUENCE 168 AA; 18838 MW; D0089CF2219C624B CRC64;
Query Match 44.4%; Score 44; DB 1; Length 168;
Best Local Similarity 35.3%; Pred. NO. 8.1;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
OY 1 KDSLOWITDQYNKSESDD 17
DB 111 KEALEWLDNDOSSAEKKD 127
RESULT 11
ID YTMK_BACSU STANDARD; PRT; 270 AA.
AC O34852;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable amino-acid ABC transporter extracellular binding protein ytmk precursor.
DN YTMK.
GN Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98048467; PubMed=9387221;
RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
RA "Sequencing and functional annotation of the Bacillus subtilis genes
RA in the 200 kb rimb-dnaB region.";
RA Microbiology 143:3431-3441(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogatawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriest R., Bouzier L., Brans A., Braum M., Brigelli S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani U.J., Connerton I.F., Cummings N.J., Daniel R.A.,

```

```

RA Raentz P., Devine K.M., Dusterhoft A., Ehlich S.D., Emerson P.T.,
RA Denizot F., Ewington J., Fabre C., Ferrari E., Foulger D.,
RA Enriac K.D., Herrington J., Kohnen S., Kohnen S., Kohnen S.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
RA Gilm S.Y., Glaser P., Goffeau A., Goldlight E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haeck J., Harwood C.R., Hentut A.,
RA Hilbert B., Holappell S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Kleier-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mamel C., Medigue C.,
RA Medina N., Melillo R.P., Mizuno M., Moestl D., Nakai S., Nodack M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presseran B., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scallan E., Schleich S., Schroeter R., Scoffone F.,
RA Seliguchi Y., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Socolini A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpetra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartot A.,
RA Vieri A., Wandut R., Wedler E., Wedler H., Wetznesger T.,
RA Winters P., Wiput A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumestein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RT Nature 390:249-256(1997).
CC -1- FUNCTION: PROBABLY PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT
CC SYSTEM YTMKLN FOR AN AMINO ACID.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (potential).
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC PROTEIN FAMILY 3.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
CC DR EMBL; AF008220; AAC00326.1; -.
CC DR EMBL; Z99119; CAB14915.1; -.
CC DR EMBL; Z99118; CAB14897.1; -.
CC DR PIR; E69996; E69996.
CC DR Subtilist; Bg13885; ytmk.
CC DR InterPro; IPR001311; SBP/glu receptor.
CC DR InterPro; IPR001638; SBP_bac_3.
CC DR Pfam; PF00497; SBP_bac_3; 1.
CC DR SMART; SM00062; PBD; 1.
CC DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
CC DR PROSITE; PS01039; SBP_BACTERIAL; 3; FALSE NEG.
CC KW Hypothetical protein; Transport; Amino-acid transport; Lipoprotein;
CC KW Membrane; Signal; Complete proteome.
CC PT SIGNAL 1 20 BY SIMILARITY.
CC FT CHAIN 21 270 PROBABLE AMINO-ACID ABC TRANSPORTER
CC FT LIPID 21 21 EXTRACELLULAR BINDING (PROBABLE).
CC SQ SEQUENCE 270 AA; 30240 MM; 6P16029F7B9C6638 CRC64;

Query Match 44.4%; Score 44; DB 1; Length 270;
Best Local Similarity 58.3%; Pred. No. 14;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

CY 3 SLOWITDOYNKE 14
DB 257 SLKMWGDDYSKE 268

RESULT 12
BIPI_TOBAC STANDARD; PRT; 290 AA.
AC Q03681;

```

```

DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Luminal binding protein 1 (Bip 1) (78 kDa glucose-regulated protein
DE homolog 1) (GRP 78-1) (Fragment).
GN Bip1
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asceridae; lamiids; Solanales; Solanaceae; Nicotiana.
OX NCBI_Taxid=4097;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92361242; PubMed=1822990;
RA Denoeck J., Goldman M.H., Demolder J., Seurinck J., Botterman J.;
RT "The tobacco luminal binding protein is encoded by a multigene
RT family."
RL Plant Cell 3:1025-1035(1991).
RN [2]
RP ERBATTM.
RA Denoeck J., Goldman M.H., Demolder J., Seurinck J., Botterman J.;
RL Plant Cell 3:1251-1251(1991).
CC -1- FUNCTION: PROBABLY PLAYS A ROLE IN FACILITATING THE ASSEMBLY OF
CC MULTIMERIC PROTEIN COMPLEXES INSIDE THE ER.
CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X60060; CAA42662.1; -
DR PIR; S21877; S21877.
DR HSSP; P08109; 1CRK.
DR InterPro; IPR000886; ER target.
DR InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; HSP70.1
DR PRINTS; PR00301; HEATSHOCK70.
DR PRODOM; PD000089; Hsp70.1.
DR PROSITE; PS00297; HSP70_1; PARTIAL.
DR PROSITE; PS00329; HSP70_2; PARTIAL.
DR PROSITE; PS01036; HSP70_3; PARTIAL.
DR PROSITE; PS00014; ER TARGET; 1.
KW ATP-binding; Endoplasmic reticulum; Multigene family.
FT NON_TER 1
FT CARBOHYD 241 241 N-LINKED (GLCNAC... ) (POTENTIAL).
FT SITE 287 290 PREVENT SECRETION FROM ER (POTENTIAL).
SQ SEQUENCE 230 AA; 32031 MW; E80F65136E64BF89 CRC64;

Query Match 44.4%; Score 44; DB 1; Length 290;
Best Local Similarity 35.3%; Pred. No. 15;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 1 KDSLOWITDQYNKESDD 17
DB 232 KEALFWLDDNOSAKED 248

RESULT 13
BIP2_TOBAC STANDARD; PRT; 292 AA.
AC 003682;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Luminal binding protein 2 (Bip 2) (78 kDa glucose-regulated protein
DE homolog 2) (GRP 78-2) (Fragment).
GN Bip2.
OS Nicotiana tabacum (Common tobacco).

```

```

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asceridae; lamiids; Solanales; Solanaceae; Nicotiana.
OX NCBI_Taxid=4097;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92361242; PubMed=1822990;
RA Denoeck J., Goldman M.H., Demolder J., Seurinck J., Botterman J.;
RT "The tobacco luminal binding protein is encoded by a multigene
RT family."
RL Plant Cell 3:1025-1035(1991).
RN [2]
RP ERBATTM.
RA Denoeck J., Goldman M.H., Demolder J., Seurinck J., Botterman J.;
RL Plant Cell 3:1251-1251(1991).
CC -1- FUNCTION: PROBABLY PLAYS A ROLE IN FACILITATING THE ASSEMBLY OF
CC MULTIMERIC PROTEIN COMPLEXES INSIDE THE ER.
CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X60059; CAA42661.1; -
DR PIR; P02662; S21878.
DR HSSP; P08109; 1CRK.
DR InterPro; IPR000886; ER target.
DR InterPro; IPR001023; Hsp70.
DR Pfam; PF00012; HSP70.1
DR PRINTS; PR00301; HEATSHOCK70.
DR PRODOM; PD000089; Hsp70.1.
DR PROSITE; PS00297; HSP70_1; PARTIAL.
DR PROSITE; PS00329; HSP70_2; PARTIAL.
DR PROSITE; PS01036; HSP70_3; PARTIAL.
DR PROSITE; PS00014; ER TARGET; 1.
KW ATP-binding; Endoplasmic reticulum; Multigene family.
FT NON_TER 1
FT CARBOHYD 241 241 N-LINKED (GLCNAC... ) (POTENTIAL).
FT SITE 289 292 PREVENT SECRETION FROM ER (POTENTIAL).
SQ SEQUENCE 232 AA; 32260 MW; 6E7A4F5107C6E2D5 CRC64;

Query Match 44.4%; Score 44; DB 1; Length 292;
Best Local Similarity 35.3%; Pred. No. 15;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 1 KDSLOWITDQYNKESDD 17
DB 232 KEALFWLDDNOSAKED 248

RESULT 14
BIP8_TOBAC STANDARD; PRT; 293 AA.
AC 003685;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Luminal binding protein 8 (Bip 8) (78 kDa glucose-regulated protein
DE homolog 8) (GRP 78-8) (Fragment).
GN Bip8.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asceridae; lamiids; Solanales; Solanaceae; Nicotiana.
OX NCBI_Taxid=4097;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92361242; PubMed=1822990;

```

RA Denoecke J., Goldman M.H., Demolder J., Seurinck J., Botterman J.,  
 RT "The tobacco luminal binding protein is encoded by a multigene  
 RT family";  
 RL Plant Cell 3:1025-1035(1991).  
 RN [2]  
 RP ERRATUM.  
 RA Denoecke J., Goldman M.H., Demolder J., Seurinck J., Botterman J.,  
 RL Plant Cell 3:1251-1251(1991).  
 CC -1- FUNCTION: PROBABLY PLAYS A ROLE IN FACILITATING THE ASSEMBLY OF  
 CC MULTIMERIC PROTEIN COMPLEXES INSIDE THE ER.  
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X60062; CAA42664.1; -  
 DR PIR; S21881; S21881.  
 DR HSP; P08109; 1CKR.  
 DR InterPro; IPR000886; ER target.  
 DR InterPro; IPR01023; Hsp70.  
 DR Pfam; PF00012; HSP70.1.  
 DR PRINTS; PR00301; HEATSHOCK70.  
 DR ProDom; PD000089; Hsp70; 1.  
 DR PROSITE; PS00297; HSP70.1; PARTIAL.  
 DR PROSITE; PS00329; HSP70.2; PARTIAL.  
 DR PROSITE; PS01036; HSP70.3; PARTIAL.  
 DR PROSITE; PS00014; ER TARGET; 1.  
 KW ATP-binding; Endoplasmic reticulum; Multigene family.  
 FT NON\_TER 1 1  
 FT CARBOHYD 244 244 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT SITE 290 293 PREVENT SECRETION FROM ER (POTENTIAL).  
 SQ SEQUENCE 293 AA; 32342 MW; 449395958456506 CRC64;  
 Query Match 44.4%; Score 44; DB 1; Length 293;  
 Best Local Similarity 35.3%; Pred. No. 15;  
 Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;  
 Oy 1 KDSLOWITDOYNKESD 17  
 Db 235 KEALMLDNDGSAEKED 251  
 RESULT 15  
 PREL STRAU STANDARD; PRT; 403 AA.  
 AC P03857;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Plasmid recombination enzyme (Mobilization protein).  
 GN PRE OR MOB.  
 OS Staphylococcus aureus.  
 OG Plasmid pE194.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 CX NCBI\_TaxId=1280;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=82167187; PubMed=6279574;  
 RA Horinouchi S., Weisblum B.;  
 RT "Nucleotide sequence and functional map of pE194, a plasmid that  
 RT specifies inducible resistance to macrolide, lincosamide, and  
 RT streptomycin type B antibiotics";  
 RL J. Bacteriol. 150:804-814(1982).  
 CC -1- FUNCTION: THE INTERACTION OF THE RSA SITE AND THE PRE PROTEIN MAY  
 CC NOT ONLY SERVES A FUNCTION IN PLASMID MAINTENANCE, BUT MAY ALSO  
 CC CONTRIBUTES TO THE DISTRIBUTION OF SMALL ANTIBIOTIC RESISTANCE  
 CC PLASMIDS AMONG GRAM-POSITIVE BACTERIA.

CC -1- MISCELLANEOUS: PRE PROTEINS CONTAIN CONSERVED POSITIVELY CHARGED  
 CC AMINO ACIDS PROBABLY INVOLVED IN THE BINDING OF THE PRE PROTEIN TO  
 CC THE RSA SITE.  
 CC -1- SIMILARITY: TO OTHER PRE PROTEINS (FROM PLASMIDS PUB110, PMW158,  
 CC PE194, PT181, PTB913), IN THEIR N-TERMINAL ONLY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; V01278; CAA4593.1; -  
 DR InterPro; IPR001668; Mob Pre.  
 DR Pfam; PF01076; Mob Pre; 1.  
 KW Plasmid; DNA-binding.  
 FT BINDING 45 45 DNA (POTENTIAL).  
 FT BINDING 115 115 DNA (POTENTIAL).  
 SQ SEQUENCE 403 AA; 47839 MW; 875867A394000FC1 CRC64;  
 Query Match 44.4%; Score 44; DB 1; Length 403;  
 Best Local Similarity 46.7%; Pred. No. 21;  
 Matches 7; Conservative 6; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 KDSLOWITDOYNKES 15  
 Db 105 KDSLEFLENEYKEN 119

Search completed: January 21, 2004, 12:07:21  
 Job time : 3.33258 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 12:04:49 ; Search time 3.25161 Seconds  
(without alignments)

532.362 Million cell updates/sec

Title: US-09-941-314-7

Perfect score: 99

Sequence: 1 KDSLOWITDOYKESDDK 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR\_76:.\*  
2: p1r1:.\*  
3: p1r2:.\*  
4: p1r3:.\*  
5: p1r4:.\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	53.5	205	2 T33117	hypothetical prote
2	47.5	48.0	809	2 E90016	hyaluronate lyase
3	47	47.5	235	1 VHVULV	nucleoprotein N -
4	47	47.5	430	2 D64151	hypothetical prote
5	47	47.5	437	2 B90451	glycosyltransferas
6	47	47.5	518	2 T00398	hypothetical prote
7	46	46.5	484	2 A89820	glutamy1-cRNA synt
8	46	46.5	525	2 E96786	protein F10A5.13 l
9	45.5	46.0	160	2 B82710	conserved hypothet
10	45	45.5	333	2 C82888	phenylalanine-trna
11	45	45.5	414	2 B81413	probable two-compo
12	45	45.5	655	2 H86222	hypothetical prote
13	44	44.4	60	2 A45681	hypothetical 7K pr
14	44	44.4	270	2 E69966	amino acid ABC tra
15	44	44.4	290	2 S21877	dnak-type molecula
16	44	44.4	292	2 S21878	dnak-type molecula
17	44	44.4	293	2 S21881	dnak-type molecula
18	44	44.4	403	1 Q05A4E	hypothetical prote
19	44	44.4	467	2 JQ0966	dnak-type molecula
20	44	44.4	474	2 E95043	hypothetical prote
21	44	44.4	478	2 G97913	phosphoglucanate d
22	44	44.4	498	2 H97214	endoglucanase, fam
23	44	44.4	663	2 T04078	dnak-type molecula
24	44	44.4	663	2 T06358	dnak-type molecula
25	44	44.4	666	2 S21879	dnak-type molecula
26	44	44.4	667	2 S21880	dnak-type molecula
27	44	44.4	668	2 T46574	dnak-type molecula
28	44	44.4	668	2 D51282	ankyrin-related pr
29	44	44.4	834	2 D51282	ankyrin-related pr

30	44	44.4	1002	2 T30546	major surface glyco
31	44	44.4	5170	2 T15148	hypothetical prote
32	44	43.4	119	2 T122029	hypothetical prote
33	43	43.4	139	2 A45361	cystatin-related e
34	43	43.4	300	2 T34213	hypothetical prote
35	43	43.4	345	2 T25863	hypothetical prote
36	43	43.4	348	2 T25863	hypothetical prote
37	43	43.4	367	1 S29871	DNase (apurinic or a
38	43	43.4	401	2 E97260	uncharacterized co
39	43	43.4	485	2 E83663	glutamy1-cRNA synt
40	43	43.4	504	2 S54744	cellulase (EC 3.2.
41	43	43.4	646	2 G85056	probable receptor-
42	43	43.4	783	2 T38891	hypothetical prote
43	43	43.4	863	2 AE0525	outer membrane ush
44	43	43.4	2285	2 T12796	probable transglyc
45	42.5	42.9	789	2 H83354	probable TonB-depe

#### ALIGNMENTS

##### RESULT 1

T33117

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jun-2000

C:Accession: T33117

R:Titin-Mollam, A.; Suterer, C.; Ozeraky, P.

submitted to the EMBL Data Library, May 1998

A:Description: The sequence of C. elegans coemid B0511.

A:Reference number: 221285

A:Accession: T33117

A:Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-205 <TIN>

A:Cross-references: EMBL:AF067608; PIDN:AACT1648.1; GSPDB:GN00019; CESP:B0511.9

A:Experimental source: strain Bristol N2; clone B0511

C:Genetics:

A:Gene: CESP:B0511.9

A:Map position: 1

A:Insertions: 69/1; 88/3; 99/3; 184/1

C:Superfamily: Caenorhabditis elegans hypothetical protein B0511.9

Query Match 53.5%; Score 53; DB 2; Length 205;

Best Local Similarity 61.5%; Pred. No. 0.97; Mismatches 2; Indels 0; Gaps 0;

Matches 8; Conservative

QY 1 KDSLOWITDOYK 13

DB 15 EDDIQLTDLQNK 27

##### RESULT 2

E90016

C:Species: Staphylococcus aureus (strain N315)

C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001

C:Accession: E90016

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc

ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramoto, K.

Lancet 357, 1255-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: E90016

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-809 <KDR>

A:Cross-references: GB:BA000018; PID:G13702002; PIDN:BAB43294.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: hyaA

	Query March	48.0%	Score 47.5;	DB 2;	Length 809;
	Best Local Similarity	47.4%	Pred. No. 31;		
	Matches	9;	Conservative	5;	Mismatches 4; Indels 1; Gaps 1.
Qy	1 KDSIQMT-TDQYNKESDDK	18			
	:   :         :				
Db	145 KDALSWLHKNAVGEKPDNK	163			

RESULT 3  
VHVLV  
nucleoprotein N - La Crosse virus  
C:Species: La Crosse virus

C:Accession: A04104  
R:Akashi, H.; Bishop, D.H.L.  
J. Virol. 45, 1155-1158, 1983  
A:Title: Comparison of the sequences and coding of Ia Grosse and snowshoe hare bunyavirus  
A:Reference number: A92989; MUID:83164355; PMID:683480  
A:Accession: A04104

A:Residues:1-235 <AKA>  
A:Cross-references: GB:K00610; NID:g210760; PIDD:AAA42782.1; PID:g210761  
C:Genetics:  
A:Map position: segment S  
C:Superfamily: bunyavirus nucleoprotein N  
C:Keywords: nucleoprotein

Query Match	47.5%	Score 47	DB 1	Length 235
Best Local Similarity	46.2%	Pred. No. 9.4		
Matches	6	Conservative	4	Mismatches 3
				Indels 0
				Gaps 0
Qy	5	QWITDQYKESDD	17	
	:	:	:	:
	:	:	:	:
	:	:	:	:
Db	101	RWVLDQYVNNDE	113	

RESULT 4  
D64151  
hypothetical protein H10404 - Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 09-Dec-2002  
C:Accession: D64151  
R:/Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weldman, J.  
; Dreyer, B.; Brandon, R.C.; FINE, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,  
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630; PMID:7542800  
A:Accession: D64151  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Releases: 1-430 <TTGR>  
A:Cross-References: GB:U32723; GB:I42023; NID:g1573363; PIDN:AAC22063.1; PID:g1573375; T  
A:Note: best homolog was a hypothetical protein from *Bacillus subtilis*  
A:Superfamily: Cell cycle protein Mes1

	Query Match	47.5%	Score 47;	DB 2;	Length 430;
	Best Local Similarity	46.7%	Pred. No. 18;		
	Matches	7;	Conservative	5;	Mismatches
				3;	Indels
				0;	Gaps
OY	1 KDSLQWITDQYNKES	15			
	:         :	7 :			
	b         :	:			
b	175 KEKLNIWTDESEEDN	189			

RESULT 5  
B90451  
glycoyltransferase, probable [imported] - Sulfolobus solfataricus  
C:Species: Sulfolobus solfataricus  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001  
C:Accession: B90451

R,She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayer, M.J.; Chan-  
Jong, I.; Jeffries, A.C.; Kozere, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P  
arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.  
submitted to GenBank, April 2001  
A:Description: *Sulfolobus solfataricus* complete genome.  
A:Reference number: A99139

A;Residues: 1-437 <KRR>  
A;Cross-references: GB:AE006641, NID:g13816086, PIDN:AAK42865.1, GSPDB:GN0015555  
C;Genetics:  
A;Gene: SS02755

Query Match	47.5%	Score 47	DB 2	Length 437
Best Local Similarity	44.4%	Pred. No. 19		
Matches	8	Conservative	5	Mismatches 5; Indels 0; Gaps 0
QY	1 KDSLQWITDYINKESDDK 18			
Db	235 RKRIQYIEDKKYQLKSDDK 252			

RESULT 6  
T00398  
hypothetical protein At2g44950 [imported] - Arabidopsis thaliana  
NAlternate names: hypothetical protein T13E15.4

C,Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 16-Feb-2001  
C,Accession: T00398; G84884  
R,Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Mason  
submitted to the EMBL Data Library, July 1997  
A,Description: *Arabidopsis thaliana* chromosome II BAC T13B15 genomic sequence.

A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-518 <ROI>  
A:Cross-references: EMBL:AC002368, NID:g3420042, PID:g2344889  
A:Experimental source: cultivar Columbia  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.  
M.; Koo, H.; Moffett, K.S.; Cronin, L.A.; Shen, M.; Vankken, S.E.; Umayam, L.; Tallon, L.; Nielsen, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter, D.  
A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: G84884  
A:Molecule type: preliminary  
A:Status: preliminary  
A:Residues: 1-518 <STO>  
A:Cross-references: GB:AE002093, NID:g4895255, PID:PAD32840.1, GSPDB:GND0139  
C:Genetic8:  
A:Gene: T13B15.4; AC2g44950  
A:Map position: 2  
A:Introns: 12/1; 85/1; 138/3; 162/3; 188/2; 253/2; 277/3; 306/3; 352/3; 420/2; 475/3  
C:Superfamily: RING finger homology  
A:462-510/Domain: RING finger homology <REN>

Query	1	KDSIQWITDQYNKESDDK	18	
db	350	EDQLRFTCTDQFKLAEDK	367	
		47.5%; Score 47; DB 2; Length 518; Best Local Similarity 44.4%; Pred. No. 23; Matches 8; Conservative 6; Mismatches 4; Indels 0; Gaps 0;		

```
RESULT 7
A89820
glutamy1-cRNA synthetase [imported] - Staphylococcus aureus (strain N315)
C1Species: Staphylococcus aureus
C1Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C1Accession: A89820
```

R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogunuma, A.; Mizutani-Oi, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hirata, K.  
Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.  
A:Reference number: A89758; MUID:21311952; PMID:11418146

A:Accession: A89820  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-484 <STR>  
A:Cross-references: GB:BA000018; PID:G13700418; PIDN:BA841716.1; GSPDB:GN00149  
A:Experimental source: strain NJ15  
A:Genetics: glx  
A:Gene: glx  
C:Superfamily: glutamate-tRNA ligase; glutamine-tRNA ligase homology

Query Match 46.5%; Score 46; DB 2; Length 484;  
Best Local Similarity 43.8%; Pred. No. 30;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 KDSLOWITDQYNKESD 16  
Db 318 KQKLAWVNNQYMKQKD 333

RESULT 8  
E96786  
protein F10A5.13 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001

C:Accession: E96786  
R:Teologio, A.; Becker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.N.; Lueros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, K.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

ker, M.; F.; Hughes, B.; Hulzar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.N.; Lueros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, K.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: E96786

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-525 <STO>

A:Cross-references: GB:AE005173; NID:G9369363; PIDN:AAF87112.1; GSPDB:GN00141

C:Genetics:

A:Gene: F10A5.13

A:Map position: 1

C:Superfamily: Arabidopsis membrane-anchored cellulase KOR

Query Match 46.5%; Score 46; DB 2; Length 525;  
Best Local Similarity 88.9%; Pred. No. 33;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KDSLOWITD 9  
Db 145 KDSLOWITD 153

RESULT 9  
E82710  
conserved hypothetical protein XF1198 [imported] - Xylella fastidiosa (strain 9asc)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000

C:Accession: E82710

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A>Note: for a complete list of authors see reference number A59328 below

A:Accession: E82710

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-160 <STM>

A:Cross-references: GB:AE003954; GB:AE003849; NID:G9106165; PIDN:AAF84008.1; GSPDB:GN001

A:Experimental source: strain 9asc

R:Simpon, A.J.G.; Reinisch, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Brienes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrex, H

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Prohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitzajima, J.P.; Krieger, J.B.; Kuramae, E.B.; Laigr

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

A:Authors: Martins, B.M.P.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri, D.A.

Rodrigues, V.; Rosa, A.U. de M.; de Rosa, V.B.; de Sa, R.G.; Santelli, R.V.; Sawasak

A:Authors: da Silva, A.C.R.; da Silva, P.R.; da Silva, F.R.; da Silva, W.A.; da Silveir

M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF1198

Query Match 46.0%; Score 45.5; DB 2; Length 160;  
Best Local Similarity 60.0%; Pred. No. 10;  
Matches 9; Conservative 2; Mismatches 1; Indels 3; Gaps 1;

Qy 5 QW1---TDQYNKESD 16  
Db 138 EWIDQVLDQYNKESD 152

RESULT 10  
C82868  
phenylalanine-tRNA synthetase alpha chain U0458 [imported] - Ureaplasma urealyticum

C:Species: Ureaplasma urealyticum

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000

C:Accession: C82868

R:Glas, J.I.; Letkowitz, E.J.; Glas, J.S.; Helner, C.R.; Chen, E.Y.; Caswell, G.H.

submitted to GenBank, February 2000

A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min

A:Reference number: A82870

A:Accession: C82868

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-333 <GLA>

A:Cross-references: GB:AE002142; GB:AF222894; NID:G6899447; PIDN:AAF30870.1; GSPDB:GN001

A:Experimental source: serovar 3; biovar 1

C:Genetics:

A:Gene: pnes; U0458

A:Genetic code: SGC3

C:Superfamily: phenylalanine-tRNA ligase alpha chain

Query Match 45.5%; Score 45; DB 2; Length 333;  
Best Local Similarity 44.4%; Pred. No. 28;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 1 KDSLOWITDQYNKESDDK 18  
Db 59 KQIEFVTDQILKEINDK 76

RESULT 11

B81413

probable two-component response regulator Cj0643 [imported] - Campylobacter jejuni (stra

C:Species: Campylobacter jejuni

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 14-Apr-2003

C:Accession: B81413

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Baham, D.; Chilling

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barr

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp

A:Reference number: A81250; MUID:20150912; PMID:10688204

A:Accession: B81413

A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-414 <PAR>  
 A:Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB75279.1; PID:g696811  
 A:Experimental source: serotype O2, strain NCTC 11168  
 C:Genetics:  
 A:Gene: Cj0643  
 C:Superfamily: response regulator with a receiver and a diguanylate cyclase (GGDEF) doma

Query Match 45.5%; Score 45; DB 2; Length 414;  
 Best Local Similarity 47.1%; Pred. No. 36;  
 Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 2 DSIQWITDYNNKESDD 18  
 DB 293 DNLQMINDEYRGVGE 309

RESULT 12  
 hypothetical protein [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
 C:Accession: H86222  
 R:Theologos, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 Ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salberg, S.L.; Schwartz, J.R.; Shih, P.; Southwick, A.M.; Sun, H.; Tallon,  
 Ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.  
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A66141; MUID:21016719; PMID:11130712  
 A:Accession: H86222  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-555 <STO>  
 A:Cross-references: GB:AE005172; NID:g2342678; PIDN:AAB70400.1; GSPDB:GM00141  
 C:Genetics:  
 A:Map position: 1  
 C:Superfamily: heat shock protein 70

Query Match 45.5%; Score 45; DB 2; Length 655;  
 Best Local Similarity 35.3%; Pred. No. 59;  
 Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 1 KDSLOWITDYNNKESDD 17  
 DB 598 KEALEMLENNNAEKED 614

RESULT 13  
 hypothetical 7K protein (gene 61.5 protein) - phage T4  
 A:Accession: A45681  
 N:Alternate names: gp 61.5  
 C:Species: phage T4  
 A:Note: host Escherichia coli  
 C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 08-Oct-1999  
 C:Accession: A45681; J05058  
 R:Sejick, H.E.; Stormo, G.D.; Dyson, R.L.; Alberts, B.M.  
 J. Virol. 67, 2303-2316, 1993  
 A:Title: Analysis of five presumptive protein-coding sequences clustered between the prf  
 A:Reference number: A45681; MUID:93188183; PMID:8383243  
 A:Accession: A45681  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-60 <SEL>  
 A:Cross-references: GB:S57514; NID:g296518; PIDN:AAB25708.1; PID:g296519  
 A:Note: sequence extracted from NCBI Backbone (NCBIN:128289; NCBI:P:128290)  
 C:Genetics:  
 A:Map position: 20.378-20.558

Query Match 44.4%; Score 44; DB 2; Length 60;  
 Best Local Similarity 46.2%; Pred. No. 5.9;  
 Matches 12; Conservative 1; Mismatches 5; Indels 8; Gaps 2;

OY 1 KDSLOWITDYNNKESDD 18  
 DB 26 KDVQWATTOYAEVVKALVKAFIDK 51

RESULT 14  
 amino acid ABC transporter (binding protein) homolog ytmk - Bacillus subtilis  
 E69996  
 C:Species: Bacillus subtilis  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
 C:Accession: E69996  
 R:Kunst, F.; Ogawara, N.; Moszer, I.; Albertin, A.M.; Alloni, G.; Azevedo, V.; Bertere,  
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho,  
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Galizzi, A.; Gallier,  
 Jech, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.  
 Koeter, P.; Koenigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
 Y. M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowaka, A.; Seror,  
 Akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpiltra, F.; Tognoni, A.; Tosato, V.; Uchiyama,  
 T.; Wintere, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida, K.  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A:Reference number: A69580; MUID:98044033; PMID:9384377  
 A:Accession: E69996  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-270 <KDN>  
 A:Cross-references: GB:Z99118; GB:Z99119; GB:AL009126; NID:g2635411; PIDN:CAB14915.1; PIDN:  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: ytmk  
 C:Superfamily: lysine-arginine-ornithine-binding protein

Query Match 44.4%; Score 44; DB 2; Length 270;  
 Best Local Similarity 58.3%; Pred. No. 31;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 3 SLOWITDYNNK 14  
 DB 257 SLKWLGGDDYSKE 268

RESULT 15  
 dnaK-type molecular chaperone b1p1 - common tobacco (fragment)  
 S21877  
 N:Alternate names: luminal binding protein b1p1  
 C:Species: Nicotiana tabacum (common tobacco)  
 C:Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 20-Aug-1999  
 C:Accession: S21877; P00264  
 R:Denoeke, J.; Goldman, M.H.; Demolder, J.; Seurinck, J.; Botterman, J.  
 submitted to the EMBL Data Library, June 1991  
 A:Description: The luminal binding protein (b1p) is encoded by a multigene family in tob  
 A:Reference number: S21877  
 A:Accession: S21877  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-290 <DEN>  
 A:Cross-references: EMBL:X60060; NID:g19804; PIDN:CAA42662.1; PID:g19805  
 A:Denoeke, J.; Goldman, M.H.S.; Demolder, J.; Seurinck, J.; Botterman, J.  
 Plant Cell 3, 1025-1035, 1991  
 A:Title: The tobacco luminal binding protein is encoded by a multigene family.  
 A:Reference number: J01360; MUID:92361242; PMID:1822990  
 A:Accession: P00264  
 A:Molecule type: mRNA

A;Residues: 263-290 <DEW>  
A;Note: translation of the nucleotide sequence is not complete  
C;Genetics:  
A;Gene: b1p1  
C;Function:  
A;Description: involved in protein folding and assembling/disassembling of protein complex  
uration steps in vivo  
C;Superfamily: heat shock protein 70  
C;Keywords: ATP; endoplasmic reticulum; molecular chaperone  
F;287-290/Region: endoplasmic reticulum retention signal

Query March 44.4%; Score 44; DB 2; Length 290;  
Best Local Similarity 35.3%; Pred. No. 34;  
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 KDSLQWITDQYNKESDD 17  
|:|:|:|:|:|:|:|:|:|  
Db 232 KEALEWLDNDNQSAREKD 248

Search completed: January 21, 2004, 12:09:17  
Job time : 4.25161 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 12:04:04 ; Search time 7.2 Seconds  
(without alignments)  
645.132 Million cell updates/sec

Title: US-09-941-314-7  
Perfect score: 99  
Sequence: 1 KDSLOWITDOYNKESDDK 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_23:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_virus:\*

16: sp\_bacteriophage:\*

17: sp\_archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99	100.0	103	4 Q8WU5	Q8WU5 homo sapien
2	99	100.0	138	4 Q8WU6	Q8WU6 homo sapien
3	99	59.6	139	11 Q8K5A3	Q8K5A3 ratuue norv
4	53	53.5	194	5 Q61819	Q61819 caenorhabd1
5	53	53.5	205	5 Q8T7Y7	Q8T7Y7 caenorhabd1
6	50	50.5	524	10 Q8SBN2	Q8SBN2 volvox cart
7	49	49.5	113	12 Q86983	Q86983 gallieria me
8	49	49.5	222	12 Q39107	Q39107 bunyavirus
9	49	49.5	2226	5 Q97225	Q97225 plasmodium
10	48	48.5	423	9 Q8SDQ4	Q8SDQ4 staphylococ
11	48	48.5	452	16 Q8NWJ3	Q8NWJ3 staphylococ
12	48	48.5	665	10 Q9M4E8	Q9M4E8 cucumis sat
13	48	48.5	668	10 Q9FSY7	Q9FSY7 corylus ave
14	48	48.5	1569	16 Q926M4	Q926M4 listeria in
15	48	48.5	1570	2 Q93R01	Q93R01 lactococcus
16	48	48.5	1844	5 Q81E31	Q81E31 plasmodium

17	47.5	48.0	809	16 Q99S67	Q99S67 staphylococ
18	47	47.5	222	12 Q39106	Q39106 bunyavirus
19	47	47.5	235	12 Q8JPR0	Q8JPR0 bunyavirus
20	47	47.5	437	17 Q97V80	Q97V80 sulfolobus
21	47	47.5	518	10 Q22157	Q22157 arabidopsis
22	47	47.5	878	10 Q8RXD6	Q8RXD6 arabidopsis
23	46	46.5	525	10 Q8LCP6	Q8LCP6 arabidopsis
24	46	46.5	525	10 Q9LR07	Q9LR07 arabidopsis
25	45.5	46.0	160	16 Q9PE30	Q9PE30 xylella fas
26	45.5	46.0	815	16 Q8NWC3	Q8NWC3 staphylococ
27	45	45.5	151	9 Q8L734	Q8L734 vibriophaga
28	45	45.5	321	2 Q8G8T6	Q8G8T6 pseudomonas
29	45	45.5	414	16 Q9PHM4	Q9PHM4 campylobact
30	45	45.5	644	16 Q8D986	Q8D986 vibrio vuln
31	45	45.5	655	10 Q04022	Q04022 arabidopsis
32	45	45.5	675	10 Q8H1B3	Q8H1B3 arabidopsis
33	45	45.5	856	10 Q9C594	Q9C594 arabidopsis
34	44	44.4	182	13 Q90ZM4	Q90ZM4 paracichthy
35	44	44.4	186	16 Q8DJP4	Q8DJP4 synechococc
36	44	44.4	279	16 Q8R5P1	Q8R5P1 fusobacteri
37	44	44.4	474	16 Q97S16	Q97S16 streptococ
38	44	44.4	474	16 Q8DR54	Q8DR54 streptococ
39	44	44.4	498	16 Q97G16	Q97G16 clostridium
40	44	44.4	570	2 Q9XDC5	Q9XDC5 streptococ
41	44	44.4	570	16 Q8NZM4	Q8NZM4 glycine max
42	44	44.4	666	10 Q39804	Q39804 glycine max
43	44	44.4	668	10 Q22639	Q22639 glycine max
44	44	44.4	675	10 Q40924	Q40924 pseudotsuga
45	44	44.4	990	5 Q8SRQ0	Q8SRQ0 encephalito

## ALIGNMENTS

### RESULT 1

Q8WU5: PRELIMINARY; PRT: 103 AA.

AC Q8WU5: 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE SC13delta.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Hamil K.G., Liu Q., Zhang Y.-L., French P.S., Hall S.H.;

RT "SC13: A novel epididymal specific member of the cystatin family.";

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF35481; AAL71992.1; -

DR InterPro; IPR00010; Cystatin.

DR Pfam; PF00031; Cystatin 1.

SQ SEQUENCE 103 AA; 12285 MW; 05DD92C47387B022 CRC64;

### Query Match

Best Local Similarity 100.0%; Score 99; DB 4; Length 103;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KDSLOWITDOYNKESDDK 18  
|||||

Db 45 KDSLOWITDOYNKESDDK 62  
|||||

### RESULT 2

Q8WU6: PRELIMINARY; PRT: 138 AA.

AC Q8WU6: 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE SC13.

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Hamil K.G., Liu Q., Zhang Y.-L., French F.S., Hall S.H.;
RT "SC13: A novel epididymal specific member of the cystatin family.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF354480; AAL71991.1; -.
DR InterPro; IPR000010; Cystatin.
DR Pfam; PF00031; Cystatin; 1.
DR SMART; SM00043; CY; 1.
SQ SEQUENCE 138 AA; 16506 MW; E49440CA3585C64 CRC64;

Query Match 100.0%; Score 99; DB 4; Length 138;
Best Local Similarity 100.0%; Pred. No. 2.1e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDSLOWITDQYNKESDDK 18
Db 45 KDSLOWITDQYNKESDDK 62

RESULT 3
08K5A3 PRELIMINARY; PRT; 139 AA.
ID 08K5A3;
AC 08K5A3;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Cystatin 11.
GN CST11.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Sprague-Dawley;
RA Hamil K.G., Hall S.H.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF501290; AAM21709.1; -.
DR InterPro; IPR000010; Cystatin.
DR Pfam; PF00031; Cystatin; 1.
DR SMART; SM00043; CY; 1.
SQ SEQUENCE 139 AA; 16686 MW; E1E36DE786B4D08C CRC64;

Query Match 59.6%; Score 59; DB 11; Length 139;
Best Local Similarity 47.1%; Pred. No. 0.24;
Matches 8; Conservative 9; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDSLOWITDQYNKESDD 17
Db 45 KETLEYVEHYNKESD 61

RESULT 4
061819 PRELIMINARY; PRT; 194 AA.
ID 061819;
AC 061819;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical 21.2 kDa protein.
GN B0511.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
SQ SEQUENCE 194 AA; 21194 MW; 30DECBABA18F263A CRC64;

Query Match 53.5%; Score 53; DB 5; Length 194;
Best Local Similarity 61.5%; Pred. No. 2.8;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 KDSLOWITDQYNK 13
Db 15 EDDIQWLTQYNK 27

RESULT 5
08T7Y7 PRELIMINARY; PRT; 205 AA.
ID 08T7Y7;
AC 08T7Y7;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical 22.5 kDa protein.
GN B0511.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RA MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RA Tin-Wollam A., Sutterer C., Ozeraky P.;
RT "The sequence of C. elegans cosmid B0511.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF067608; AAC17648.2; -.
DR WormPep; B0511.9b; CE17348.
KW Hypothetical protein.
SQ SEQUENCE 205 AA; 22520 MW; 96036895513A34A7 CRC64;

Query Match 53.5%; Score 53; DB 5; Length 205;
Best Local Similarity 61.5%; Pred. No. 3;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 KDSLOWITDQYNK 13

```

```

RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RA Tin-Wollam A., Sutterer C., Ozeraky P.;
RT "The sequence of C. elegans cosmid B0511.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF067608; AAC17648.2; -.
DR WormPep; B0511.9a; CE26854.
KW Hypothetical protein.
SQ SEQUENCE 194 AA; 21194 MW; 30DECBABA18F263A CRC64;

Query Match 53.5%; Score 53; DB 5; Length 194;
Best Local Similarity 61.5%; Pred. No. 2.8;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 KDSLOWITDQYNK 13
Db 15 EDDIQWLTQYNK 27

RESULT 5
08T7Y7 PRELIMINARY; PRT; 205 AA.
ID 08T7Y7;
AC 08T7Y7;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical 22.5 kDa protein.
GN B0511.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RA MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RA Tin-Wollam A., Sutterer C., Ozeraky P.;
RT "The sequence of C. elegans cosmid B0511.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF067608; AAC15538.1; -.
DR WormPep; B0511.9b; CE17348.
KW Hypothetical protein.
SQ SEQUENCE 205 AA; 22520 MW; 96036895513A34A7 CRC64;

Query Match 53.5%; Score 53; DB 5; Length 205;
Best Local Similarity 61.5%; Pred. No. 3;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 KDSLOWITDQYNK 13

```

Db 15 RDDQWLTDLQNK 27

## RESULT 6

Q9SBN2 PRELIMINARY; PRT; 524 AA.  
AC Q9SBN2; 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
RT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Protein disulfide isomerase.  
GN PDI.  
OS Volvox carteri f. nagariensis.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
OC Volvocaceae; Volvox.  
OX NCBI\_TaxID=3068;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20119586; PubMed=10654090;  
RA Meisner M., Stark K., Greenar B., Kirk D.L., Schmitt R.;  
RT "Volvox germline-specific genes that are putative targets of RegA  
repression encode chloroplast proteins.";  
RL Curr. Genet. 36:363-370(1999).  
DR HSSP; P07237; IMK.  
DR InterPro; IPR005792; Disulphide\_isom.  
DR InterPro; IPR005788; Disulph\_isom.  
DR InterPro; IPR000886; ER target.  
DR InterPro; IPR006662; Thiorod.  
DR InterPro; IPR006662; Thiorodox\_dom2.  
DR Pfam; PF00085; Thiorodox\_2.  
DR PRINTS; PRO0421; THIOREDOXIN.  
DR TIGRFAMs; TIGR01130; ER\_PDI\_fam; 1.  
DR TIGRFAMs; TIGR01126; pdi\_dom; 2.  
DR PROSITE; PS00014; ER\_TARGET; 1.  
DR PROSITE; PS00194; THIOREDOXIN; 2.  
KW Isomerase; Redox-active center.  
SQ SEQUENCE 524 AA; 57331 MW; CF7770D3D1C291AA CRC64;

Query Match 50.5%; Score 50; DB 10; Length 524;  
Best Local Similarity 38.9%; Pred. No. 23;  
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Cy 1 KDSLOWITDQYNKESDDK 18  
Db 22 KHQLAMASDEYEDDDDE 39

## RESULT 7

Q86983 PRELIMINARY; PRT; 113 AA.  
AC Q86983; 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE Orf1.  
GN ORF1.  
OS Gallieria mellonella nuclear polyhedrosis virus (GmNPV).  
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
OC Nucleopolyhedrovirus.  
OX NCBI\_TaxID=10447;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TRANSPOSON=hitchhiker;  
RX MEDLINE=96187810; PubMed=8614994;  
RA Bauser C.A., Elick T.A., Frazer W.J.;  
RT "Characterization of hitchhiker, a transposon insertion frequently  
RT associated with baculovirus FP mutants derived upon passage in the TN-  
RT 368 cell line.";  
RL Virology 216:235-237(1996).  
DR EMBL; S81557; AAB36373.2;  
SQ SEQUENCE 113 AA; 13641 MW; F68B548415F229A9 CRC64;

Query Match 49.5%; Score 49; DB 12; Length 113;  
Best Local Similarity 56.2%; Pred. No. 6.4;  
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Cy 1 KDSLOWITDQYNKESD 16  
Db 23 RDSQWLTDLQYSCMD 38

## RESULT 8

O39107 PRELIMINARY; PRT; 222 AA.  
AC O39107; 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE Nucleocapsid protein (Fragment).  
OS Bunyavirus La Crosse.  
OC Viruses; ssRNA negative-strand viruses; Bunyaviridae; Orthobunyavirus.  
OX NCBI\_TaxID=11577;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=prototype;  
RX MEDLINE=98449861; PubMed=9774588;  
RA Chandler L.J., Borucki M.K., Dobie D.K., Wasielecki L.P.,  
RA Thompson W.H., Gundersen C.B., Case K., Beatty B.J.;  
RT "Characterization of La Crosse virus RNA in autopsied central nervous  
RT system tissues.";  
RL J. Clin. Microbiol. 36:3332-3336(1998).  
DR EMBL; AF025479; AAB81525.1;  
DR InterPro; IPR001784; Bunya\_nucleocap.  
DR Pfam; PF00952; Bunya\_nucleocap; 1.  
DR ProDom; PD001909; Bunya\_nucleocap; 1.  
FT NON\_TER 1 1  
FT NON\_TER 222 222  
SQ SEQUENCE 222 AA; 25045 MW; 6CC80D2C9234C609 CRC64;

Query Match 49.5%; Score 49; DB 12; Length 222;  
Best Local Similarity 46.2%; Pred. No. 13;  
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Cy 5 QWITDQYNKESDD 17  
Db 96 RWVLDQYNEDDDE 108

## RESULT 9

O97225 PRELIMINARY; PRT; 2226 AA.  
AC O97225; 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Hypothetical protein, conserved.  
GN MAJ3P2.2.  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99376085; PubMed=10448855;  
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,  
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltrielli T.,  
RA Gentles S., Gilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,  
RA Horrocks P., Jagels K., Jessal B., Kyes S., McLean J., Moule S.,  
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,  
RA Rutter S., Skellon J., Squares R., Squares S., Sulston J.B.,  
RA Whitehead S., Woodward J.R., Newbold C., Barrett B.G.;  
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium  
RT falciparum.";  
RL Nature 400:532-538(1999).  
RN [2]

```

GN      MW1405.  Staphylococcus aureus (strain MW2) .
OC      Bacteria; Firmicutes; Bacillales; Staphylococcus.
CX      NCBI_TaxID=196620;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=22040711; PubMed=12044378;
RA      Baba Y., Takuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA      Nagai Y., Iwama N., Asano K., Naito T., Kuroda H., Cui L.,
RA      Yamamoto K., Hiramatsu K.;
RT      "Genome and virulence determinants of high virulence community-
RT      acquired MRSA."
RL      Lancet 359:1819-1827(2002) .
DR      EMBL; AP004827; BAB95270.1; -
DR      InterPro; IPR000330; SNF2_N.
KW      Pfam; PF00176; SNF2_N; 1.
SQ      SEQUENCE 452 AA; 52770 MW; C5ED88ADF4BFD192 CRC64;

Query Match      48.5%; Score 48; DB 16; Length 452;
Best Local Similarity 50.0%; Pred. No. 40;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0

QY      1 KDSLOWITDOYKNE 14
      ::::|::|
Db      106 KENTKWLCDYKKE 119

RESULT 12
O9M4E8      PRELIMINARY; PRT; 665 AA.
AC      O9M4E8.
DT      01-OCT-2000 (TREMBLrel.15, Created)
DT      01-OCT-2000 (TREMBLrel.15, Last sequence update)
DT      01-MAR-2003 (TREMBLrel.23, Last annotation update)
DE      Heat shock protein 70.
OS      Cucumis sativus (Cucumber).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosids I; Cucurbitales; Cucurbitaceae; Cucumbrs.
OX      NCBI_TaxID=3659;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=20117683; PubMed=10651811;
RA      Kindl H., Diefenbach U.;
RT      "The membrane-bound DnaJ protein located at the cytosolic site of
RT      glyoxysomes specifically binds the cytosolic isoform 1 of Hsp70 but
RT      not other Hsp70 species."
RL      Eur. J. Biochem. 267:746-754(2000) .
CC      -1- FUNCTION: PROBABLY PLAYS A ROLE IN FACILITATING THE ASSEMBLY OF
CC      MULTIMERIC PROTEIN COMPLEXES INSIDE THE ER (BY SIMILARITY) .
CC      -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN (BY SIMILARITY) .
CC      -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
DR      EMBL; AJ249329; CAB72128.1; -
DR      HSSP; P19120; JHSC.
DR      InterPro; IPR000886; ER_target.
DR      InterPro; IPR01023; Hsp70.
DR      Pfam; PF00012; HSP70; 1.
DR      PRINTS; PR00301; HEATSHOCK70.
DR      ProDom; PD000089; Hsp70; 1.
DR      PROSITE; PS00014; ER_TARGET; 1.
DR      PROSITE; PS00297; HSP70_1; 1.
DR      PROSITE; PS00329; HSP70_2; 1.
DR      PROSITE; PS01036; HSP70_3; 1.
KW      ATP-binding; Endoplasmic reticulum.
SQ      SEQUENCE 665 AA; 73473 MW; 0A430B71804D6186 CRC64;

Query Match      48.5%; Score 48; DB 10; Length 665;
Best Local Similarity 41.2%; Pred. No. 60;
Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0

QY      1 KDSLOWITDOYKNE 17
      ::::|::|

```

Db 608 KDALEMLDNDQSAEKED 624

## RESULT 13

Q9FSY7 PRELIMINARY; PRT; 668 AA.

AC Q9FSY7; 01-MAR-2001 (TRENBLREL. 16, Last sequence update)

DT 01-MAR-2001 (TRENBLREL. 16, Last sequence update)

DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)

DE Putative luminal binding protein.

GN Bip.

OS Corylus avellana (European hazel).

OC Burkholderia viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucotyledons I; Fagales; Betulaceae; Corylus.

OK NCBI\_TaxID=13451;

RN [1]

RP SEQUENCE FROM N.A.

RA Gruen S., Suphiglu C., Volkman D.,

RT "Molecular cloning and characterization of a novel hazel pollen

allergen identified as a luminal binding protein (Bip)."

RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.

CC -1- FUNCTION: PROBABLY PLAYS A ROLE IN FACILITATING THE ASSEMBLY OF

MULTIMERIC PROTEIN COMPLEXES INSIDE THE ER (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: ENDOPLASMIC RETICULUM LUMEN (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.

DR EMBL; AJ295617; CAC14168.1; -.

DR HSPSP; P19120; HSC.

DR InterPro; IPR000886; ER\_target.

DR InterPro; IPR001023; Hsp70.

DR Pfam; PF00012; HSP70; 1.

DR PRINTS; PR00301; HEATSHOCK70.

DR PRODOM; PD000089; HSP70; 1.

DR PROSITE; PS00014; ER\_TARGET; 1.

DR PROSITE; PS00297; HSP70\_1; 1.

DR PROSITE; PS00329; HSP70\_2; 1.

DR PROSITE; PS01036; HSP70\_3; 1.

DR ATP-binding; Endoplasmic reticulum.

KW SEQUENCE 668 AA; 73564 MW; 8243BA3FC9C10D9 CRC64;

Query Match 48.5%; Score 48; DB 10; Length 668;

Best Local Similarity 41.2%; Pred. No. 61; Matches 7; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 KDSLOWITDQYNKESDD 17

Db 608 KDALEMLDNDQSAEKED 624

## RESULT 14

Q926N4 PRELIMINARY; PRT; 1569 AA.

AC Q926N4; 01-DEC-2001 (TRENBLREL. 19, Created)

DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)

DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)

DE Hypothetical protein p110021.

GN p110021.

OS Listeria innocua.

OC Plasmid p1100.

OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.

OK NCBI\_TaxID=1642;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=CLIP 11262 / Serovar 6a;

RA PubMed=11679669;

RA Glaser P., Frangul L., Buchrieser C., Rusnok C., Amend A.,

Baquerio F., Berche P., Bloeker H., Brandt P., Chatriotory T.,

Charbit A., Chetouani F., Couve E., de Darvar A., Dehoux P.,

Domnan B., Dominguez-Bernal G., Duchaud E., Durant L., Dussauget O.,

Entian K.-D., Fajhi H., Garcia-del Portillo F., Garrido P.,

Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,

RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurupkat G.,

Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,

RA Nordstjek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,

RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,

RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,

RT "Comparative genomics of Listeria species."

RL Science 294:849-852(2001).

DR EMBL; AL592102; CAC42019.1; -.

DR InterPro; IPR001410; DEND.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR002296; N12N6\_mltfrase.

DR Pfam; PF00271; Helicase\_C; 1.

DR PRINTS; PR00507; N12N6MTFRASE.

DR SMART; SM00487; DEXDC; 1.

DR SMART; SM00490; HELIC\_C; 1.

KW Plasmid; Hypothetical protein; Complete proteome.

KW SEQUENCE 1569 AA; 178613 MW; 0654EC32B84B6C04 CRC64;

Query Match 48.5%; Score 48; DB 16; Length 1569;

Best Local Similarity 38.9%; Pred. No. 1.5e+02; Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 KDSLOWITDQYNKESDDK 18

Db 1507 RPALEMITDQYQVKTDK 1524

## RESULT 15

Q93R01 PRELIMINARY; PRT; 1570 AA.

AC Q93R01; 01-DEC-2001 (TRENBLREL. 19, Created)

DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)

DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)

DE Endonuclease and methylase Llagi.

OS Lactococcus lactis.

OC Plasmid pEM104.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.

OK NCBI\_TaxID=1358;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=M10;

RA MEDLINE=21303253; PubMed=11410355;

RA "The Llagi restriction and modification system of Lactococcus lactis

RT W10 consists of only one single polypeptide."

RT FEMS Microbiol. Lett. 200:91-96(2001).

DR EMBL; AF097471; AAK71920.1; -.

DR InterPro; IPR001410; DEND.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR002296; N12N6\_mltfrase.

DR Pfam; PF00271; Helicase\_C; 1.

DR PRINTS; PR00507; N12N6MTFRASE.

DR SMART; SM00487; DEXDC; 1.

DR SMART; SM00490; HELIC\_C; 1.

DR PROSITE; PS00092; N6\_MTASE; 1.

DR ATP-binding; Helicase; Hydrolase; Methyltransferase; Plasmid.

KW SEQUENCE 1570 AA; 179355 MW; 141904EFC672793C CRC64;

Query Match 48.5%; Score 48; DB 2; Length 1570;

Best Local Similarity 38.9%; Pred. No. 1.5e+02;

Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 KDSLOWITDQYNKESDDK 18

Db 1508 RPALEMITDQYQVKTDK 1525

Search completed: January 21, 2004, 12:08:38

Job time : 9.2 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: January 21, 2004, 12:05:24 ; Search time 3.36774 Seconds  
(without alignments)  
226.144 Million cell updates/sec

Title: US-09-941-314-7

Perfect score: 99  
Sequence: 1 KDSIQWITTDQYNKESDDK 18Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

Issued Patents AA: \*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/1/1aa/PTCUS.COMB.pep: \*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	47.5	142	US-09-431-480-4	Sequence 4, Appli
2	47	47.5	142	US-09-617-302-4	Sequence 4, Appli
3	47	47.5	264	US-09-107-532A-4481	Sequence 4, Appli
4	46	46.5	484	US-08-913-578-2	Sequence 2, Appli
5	46	46.5	484	US-08-785-427-2	Sequence 2, Appli
6	44	44.4	474	US-09-058-692-2	Sequence 2, Appli
7	44	44.4	474	US-09-584-628-2	Sequence 2, Appli
8	44	44.4	655	US-09-632-538C-36	Sequence 36, Appli
9	43	43.4	142	US-09-431-480-3	Sequence 3, Appli
10	43	43.4	142	US-09-617-302-3	Sequence 3, Appli
11	43	43.4	2188	US-09-328-352-7763	Sequence 7763, Ap
12	43	43.4	2285	US-09-308-375-2	Sequence 2, Appli
13	42.5	42.9	708	US-09-252-991A-25303	Sequence 25303, A
14	42	42.4	222	US-08-556-965-2	Sequence 2, Appli
15	42	42.4	222	5472691-3	Patent No. 5472691
16	42	42.4	240	US-08-023-980B-45	Sequence 45, Appli
17	42	42.4	240	US-08-486-953A-53	Sequence 53, Appli
18	42	42.4	240	US-08-679-493A-186	Sequence 186, App
19	42	42.4	240	5472691-2	Patent No. 5472691
20	42	42.4	363	US-09-464-035A-3	Sequence 3, Appli
21	42	42.4	363	US-09-464-035A-5	Sequence 5, Appli
22	42	42.4	371	US-09-328-352-6825	Sequence 6825, Ap
23	42	42.4	404	US-09-328-352-8182	Sequence 8182, Ap
24	42	42.4	623	US-09-107-532A-4726	Sequence 4726, Ap
25	42	42.4	976	US-09-104-324B-4	Sequence 4, Appli
26	41	41.4	443	US-09-134-001C-3183	Sequence 3183, Ap
27	41	41.4	460	US-09-174-768-4	Sequence 4, Appli

28	41	41.4	888	US-09-134-001C-3032	Sequence 3032, Ap
29	40	40.4	302	US-09-107-532A-5280	Sequence 5280, Ap
30	40	40.4	501	US-09-134-001C-4115	Sequence 4115, Ap
31	40	40.4	1114	US-08-811-583-2	Sequence 2, Appli
32	40	40.4	1177	US-09-795-927-2	Sequence 2, Appli
33	40	40.4	1844	US-08-851-567B-53	Sequence 53, Appli
34	40	40.4	2504	US-08-851-567B-12	Sequence 12, Appli
35	39.5	39.9	125	US-08-408-095-15	Sequence 35, Appli
36	39.5	39.9	439	US-09-507-765-33	Sequence 33, Appli
37	39	39.4	72	US-09-328-352-5400	Sequence 3400, Ap
38	39	39.4	111	US-09-775-932-20	Sequence 20, Appli
39	39	39.4	111	US-09-775-932-22	Sequence 22, Appli
40	39	39.4	116	US-09-775-932-16	Sequence 16, Appli
41	39	39.4	139	US-08-791-532-4	Sequence 4, Appli
42	39	39.4	139	US-09-314-777-4	Sequence 4, Appli
43	39	39.4	174	US-09-328-869-17	Sequence 17, Appli
44	39	39.4	174	US-09-629-774A-17	Sequence 17, Appli
45	39	39.4	190	US-09-328-352-7389	Sequence 7389, Ap

## ALIGNMENTS

```
RESULT 1
US-09-431-480-4
Sequence 4, Application US/09431480
Patent No. 6235708
GENERAL INFORMATION:
APPLICANT: Holloway, James L.
TITLE OF INVENTION: TESTS SPECIFIC CYSTATIN-LIKE PROTEIN CYSTATIN T
FILE REFERENCE: 98-72
CURRENT APPLICATION NUMBER: US/09/431,480
EARLIER FILING DATE: 1999-11-01
EARLIER APPLICATION NUMBER: 60/109,217
EARLIER FILING DATE: 1998-11-20
EARLIER APPLICATION NUMBER: 60/156,382
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 142
TYPE: PRT
ORGANISM: Homo sapiens
US-09-431-480-4

Query Match          47.5%  Score 47;  DB 3;  Length 142;
Best Local Similarity 50.0%;  Pred. No. 4.5;
Matches 9;  Conservative 3;  Mismatches 6;  Indels 0;  Gaps 0;

Qy      1 KDSIQWITTDQYNKESDDK 18
Db      46 KQCLWFMQBYNKESDDK 63

RESULT 2
US-09-617-302-4
Sequence 4, Application US/09617302
Patent No. 6245529
GENERAL INFORMATION:
APPLICANT: Holloway, James L.
TITLE OF INVENTION: TESTS SPECIFIC CYSTATIN-LIKE PROTEIN CYSTATIN T
FILE REFERENCE: 98-72 C1
CURRENT APPLICATION NUMBER: US/09/617,302
CURRENT FILING DATE: 2000-07-17
PRIOR APPLICATION NUMBER: 09/431,480
PRIOR FILING DATE: 1999-11-01
PRIOR APPLICATION NUMBER: 60/109,217
PRIOR FILING DATE: 1998-11-20
PRIOR APPLICATION NUMBER: 60/156,382
PRIOR FILING DATE: 1999-09-28
NUMBER OF SEQ ID NOS: 22
```

SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO: 4  
LENGTH: 142  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-617-302-4

Query Match 47.5%; Score 47; DB 3; Length 142;  
Best Local Similarity 50.0%; Pred. No. 4.5;  
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 KDSLOWITDQYNKESDDK 18  
DB 46 KQCLWFMQGEYVKESEDK 63

RESULT 3  
US-09-107-532A-4481  
Sequence 4481, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bueh  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Arinello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 4481:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 264 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) LOCATION 1...264  
SEQUENCE DESCRIPTION: SEQ ID NO: 4481:  
US-09-107-532A-4481  
Query Match 47.5%; Score 47; DB 4; Length 264;  
Best Local Similarity 38.9%; Pred. No. 9.1;  
Matches 7; Conservative 5; Mismatches 6; Indels 0; Gaps 0;  
QY 1 KDSLOWITDQYNKESDDK 18  
DB 46 KQCLWFMQGEYVKESEDK 63

DB 247 KRGVWLEDDQYRETEK 264

RESULT 4  
US-08-913-578-2  
Sequence 2, Application US/08913578  
Patent No. 6218159  
GENERAL INFORMATION:  
APPLICANT: Hodgson, John  
APPLICANT: Lawlor, Elizabeth  
TITLE OF INVENTION: NO. 6218159el tRNA synthetase  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/913,578  
FILING DATE: 17-JAN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9601069.9  
FILING DATE: 19-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimm, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P31352  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 484 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-913-578-2  
Query Match 46.5%; Score 46; DB 3; Length 484;  
Best Local Similarity 43.8%; Pred. No. 25;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
QY 1 KDSLOWITDQYNKESD 16  
DB 318 KQCLWVNNQYMKQKD 333  
RESULT 5  
US-08-785-427-2  
Sequence 2, Application US/08785427  
Patent No. 6238900  
GENERAL INFORMATION:  
APPLICANT: Hodgson, John  
APPLICANT: Lawlor, Elizabeth  
TITLE OF INVENTION: NO. 6238900el tRNA synthetase  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/785,427  
FILING DATE: 17-JAN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 9601069.9  
FILING DATE: 19-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimml, Edward R.  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: Pj1352  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 484 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-785-427-2

Query Match 46.5%; Score 46; DB 3; Length 484;  
Best Local Similarity 43.8%; Pred. No. 25;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 KDSLOWITDQYNKESD 16  
Db 318 KQKLAWNQNMKQKD 333

RESULT 6  
US-09-058-692-2  
Sequence 2, Application US/09058692  
Patent No. 6162618  
GENERAL INFORMATION:  
APPLICANT: Warren, Richard L.  
TITLE OF INVENTION: 6-phosphogluconate dehydr  
TITLE OF INVENTION: Ogenase  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert, Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-2793  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/058,692  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Falk, Stephen T  
REGISTRATION NUMBER: 36,795  
REFERENCE/DOCKET NUMBER: GM10140  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-994-2488  
TELEFAX: 215-994-2222

TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 474 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-058-692-2

Query Match 44.4%; Score 44; DB 3; Length 474;  
Best Local Similarity 61.5%; Pred. No. 50;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 LOWITDQYNKESD 16  
Db 375 LQKITDAYNRDAD 387

RESULT 7  
US-09-584-628-2  
Sequence 2, Application US/09584628  
Patent No. 6309866  
GENERAL INFORMATION:  
APPLICANT: Warren, Richard L.  
TITLE OF INVENTION: 6-phosphogluconate dehydr  
TITLE OF INVENTION: Ogenase  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dechert, Price & Rhoads  
STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103-2793  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/584,628  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/058,692  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Falk, Stephen T  
REGISTRATION NUMBER: 36,795  
REFERENCE/DOCKET NUMBER: GM10140  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-994-2488  
TELEFAX: 215-994-2222  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 474 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-584-628-2

Query Match 44.4%; Score 44; DB 4; Length 474;  
Best Local Similarity 61.5%; Pred. No. 50;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 4 LOWITDQYNKESD 16  
Db 375 LQKITDAYNRDAD 387

RESULT 8  
US-09-632-538C-36

```
; Sequence 36, Application US/09632538C
; Patent No. 6440674
; GENERAL INFORMATION:
; APPLICANT: Miera, Santosh et al.
; TITLE OF INVENTION: PLANT PROMOTER DERIVED FROM LUMINAL BINDING PROTEIN GENE AND METH
; FILE REFERENCE: 54359
; CURRENT APPLICATION NUMBER: US/09/632,538C
; CURRENT FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 36
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Pseudocarya menziesii
US-09-632-538C-36

Query Match          44.4%; Score 44; DB 4; Length 655;
Best Local Similarity 35.3%; Pred. No. 72;
Matches 6; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy      1 KDSLOWITDQYNKESDD 17
       | : | : | : | : | : |
Db      598 KEALEWLDNDQSAERD 614

RESULT 9
US-09-431-480-3
; Sequence 3, Application US/09431480
; Patent No. 6235708
; GENERAL INFORMATION:
; APPLICANT: Holloway, James L.
; TITLE OF INVENTION: TESTIS SPECIFIC CYSTATIN-LIKE PROTEIN CYSTATIN T
; FILE REFERENCE: 98-72
; CURRENT APPLICATION NUMBER: US/09/431,480
; CURRENT FILING DATE: 1999-11-01
; EARLIER APPLICATION NUMBER: 60/109,217
; EARLIER FILING DATE: 1998-11-20
; EARLIER APPLICATION NUMBER: 60/156,382
; EARLIER FILING DATE: 1999-09-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-431-480-3

Query Match          43.4%; Score 43; DB 3; Length 142;
Best Local Similarity 44.4%; Pred. No. 19;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy      1 KDSLOWITDQYNKESDD 18
       | : | : | : | : | : |
Db      46 KQCVWFANKEYNKESDD 63

RESULT 10
US-09-617-302-3
; Sequence 3, Application US/09617302
; Patent No. 6245529
; GENERAL INFORMATION:
; APPLICANT: Holloway, James L.
; APPLICANT: Feldhaus, Andrew
; TITLE OF INVENTION: TESTIS SPECIFIC CYSTATIN-LIKE PROTEIN CYSTATIN T
; FILE REFERENCE: 98-72 C1
; CURRENT APPLICATION NUMBER: US/09/617,302
; CURRENT FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/431,480
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: 60/109,217
; PRIOR FILING DATE: 1998-11-20
```

```
; PRIOR APPLICATION NUMBER: 60/156,382
; PRIOR FILING DATE: 1999-09-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-617-302-3

Query Match          43.4%; Score 43; DB 3; Length 142;
Best Local Similarity 44.4%; Pred. No. 19;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy      1 KDSLOWITDQYNKESDD 18
       | : | : | : | : | : |
Db      46 KQCVWFANKEYNKESDD 63

RESULT 11
US-09-328-352-7763
; Sequence 7763, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7763
; LENGTH: 2188
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7763

Query Match          43.4%; Score 43; DB 4; Length 2188;
Best Local Similarity 50.0%; Pred. No. 3.9e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy      1 KDSLOWITDQYNKESD 16
       | | | | | : | : | : |
Db      1166 KLSLDWEAKYGRDSD 1181

RESULT 12
US-09-308-375-2
; Sequence 2, Application US/09308375
; Patent No. 6300117
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394-PCT
; CURRENT APPLICATION NUMBER: US/09/308,375
; CURRENT FILING DATE: 1999-05-14
; EARLIER APPLICATION NUMBER: EP9719636.4
; EARLIER FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-308-375-2

Query Match          43.4%; Score 43; DB 4; Length 2285;
Best Local Similarity 52.9%; Pred. No. 4.1e+02;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy      1 KDSLOWITDQYNKESDD 17
       | : | : | : | : | : |
Db      1995 QDSIQKLTQINQYSLD 2011
```

RESULT 13  
US-09-252-991A-25303  
; Sequence 25303, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 25303  
; LENGTH: 708  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-25303

Query Match 42.9%; Score 42.5; DB 4; Length 708;  
Best Local Similarity 60.0%; Pred. No. 1.3e+02;  
Matches 9; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 3 SLOWITDOYNKESD 17  
||:|||||:|  
Db 275 SLRWIGD-YNEEDSD 288

RESULT 14  
US-08-556-965-2  
; Sequence 2, Application US/08556965  
; Patent No. 6025540  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: Production of EC-SOD  
; NUMBER OF SEQUENCES: 31  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25 (BPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/556,965  
; FILING DATE:  
; CLASSIFICATION: 800  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 222 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-556-965-2

Query Match 42.4%; Score 42; DB 3; Length 222;  
Best Local Similarity 46.7%; Pred. No. 44;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 DSLQWITDOYNKESD 16  
||:|||||:|  
Db 12 DSAEWIRDMYAKVTE 26

RESULT 15  
5472691-3  
; Patent No. 5472691  
; APPLICANT: MARKLUND, STEFAN; EDLUND, THOMAS  
; TITLE OF INVENTION: SUPEROXIDE DISMUTASE  
; NUMBER OF SEQUENCES: 7

CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/125,744  
; FILING DATE: 24-Sep-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 897,624  
; FILING DATE: 12-JUN-1992  
; APPLICATION NUMBER: 576,114  
; FILING DATE: 27-Aug-1990  
; APPLICATION NUMBER: 902,596  
; FILING DATE: 02-Sep-1986  
; SEQ ID NO: 3;  
; LENGTH: 222  
5472691-3

Query Match 42.4%; Score 42; DB 6; Length 222;  
Best Local Similarity 46.7%; Pred. No. 44;  
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 DSLQWITDOYNKESD 16  
||:|||||:|  
Db 12 DSAEWIRDMYAKVTE 26

Search completed: January 21, 2004, 12:09:58  
Job time : 4.36774 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 12:16:36 ; Search time 21 Seconds  
(without alignments)  
82.430 Million cell updates/sec

Title: US-09-941-314-7

Perfect score: 99  
Sequence: 1 KDSLQWITPDQYNKESDDK 18

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 3212

Minimum DB seq length: 0  
Maximum DB seq length: 18

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.76:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25.5	25.8	14	2	S68095
2	25	25.3	13	2	S66235
3	25	25.3	15	2	A60156
4	23	23.2	15	2	S57584
5	23	23.2	15	2	G35141
6	23	23.2	18	2	SS2125
7	22	22.2	15	2	P00175
8	22	22.2	17	2	C85956
9	22	22.2	18	2	A39040
10	22	22.2	18	2	PX0081
11	21.5	21.7	15	2	B60763
12	21	21.2	10	1	RHLMS
13	21	21.2	14	2	PC1215
14	21	21.2	14	2	B83836
15	21	21.2	15	2	S21238
16	21	21.2	15	2	PT0205
17	20	20.2	5	2	A60803
18	20	20.2	10	2	PT0289
19	20	20.2	12	2	A40763
20	20	20.2	13	2	S14316
21	20	20.2	14	2	S29209
22	20	20.2	14	2	S29789
23	20	20.2	15	2	P00174
24	20	20.2	15	2	G49655
25	20	20.2	15	2	PD0444
26	20	20.2	16	2	H29501
27	20	20.2	16	2	B45895
28	20	20.2	16	2	P54226
29	20	20.2	17	2	S24570

30	20	20.2	17	2	A60570	Ig mu heavy chain
31	19.5	19.7	17	2	S51736	T-cell receptor be
32	19	19.2	8	2	A61328	trypsin (EC 3.4.21
33	19	19.2	10	2	PT0213	T-cell receptor al
34	19	19.2	11	2	PH0904	T-cell receptor be
35	19	19.2	12	2	PT0228	Ig heavy chain CDR
36	19	19.2	12	2	PT0274	Ig heavy chain CRD
37	19	19.2	14	2	A28018	very late antigen-
38	19	19.2	14	2	S66234	sperm motility inh
39	19	19.2	16	2	S22040	cob protein - comm
40	19	19.2	17	2	A296834	tip leader peptide
41	19	19.2	17	2	B28027	protein P4 - curle
42	19	19.2	18	2	S29491	GMP-binding protei
43	19	19.2	18	2	A56871	retinol-binding pr
44	18	18.2	6	2	PT0533	T-cell receptor be
45	18	18.2	7	2	PT0576	T-cell receptor be

#### ALIGNMENTS

RESULT 1  
S68095  
Calcium-binding protein, 23K - Orchestia cavimana (fragment)  
C/Species: Orchestia cavimana  
C/Date: 21-Apr-1997 #sequence\_revision 21-Apr-1997 #text\_change 29-Aug-1997  
R/Luquet, G.; Testeniere, O.; Graf, F.  
Biochim. Biophys. Acta 1293, 272-276, 1996  
A/Title: Characterization and N-terminal sequencing of a calcium binding protein from th  
A/Reference number: S68095; MUID:96202045; PMID:8620040  
A/Accession: S68095  
A/Molecule type: protein  
A/Residues: 1-14 <LUG>  
C/Keywords: calcium binding

Query Match 25.8%; Score 25.5; DB 2; Length 14;  
Best Local Similarity 46.2%; Pred. No. 8e+02; 2; Mismatches 3; Gaps 1;  
Matches 6; Conservative 2; Indels 3; Gaps 1;

QY 6 WITPDQYNKESDDK 18  
DB 1 WDDDE---ESDDR 10

RESULT 2  
S66235  
sperm motility inhibitor protein - pig (fragment)  
C/Species: Sus scrofa domestica (domestic pig)  
C/Date: 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 13-Mar-1997  
R/Iwamoto, T.; Hiroaki, H.; Furutachi, Y.; Wada, K.; Satoh, M.; Satoh, M.; Osada, T.; Gag  
FEBS Lett. 368, 420-424, 1995  
A/Title: Cloning of boar SMI gene which is expressed specifically in seminal vesicle an  
A/Reference number: S66233; MUID:95361914; PMID:7655150  
A/Accession: S66235  
A/Status: preliminary  
A/Molecule type: protein  
A/Residues: 1-13 <IWA>  
A/Note: Pro-6 was also found

Query Match 25.3%; Score 25; DB 2; Length 13;  
Best Local Similarity 83.3%; Pred. No. 8e+02;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 NKESDD 17  
DB 3 NKESDD 8

RESULT 3  
A60156  
cellulase (EC 3.2.1.4) - Clostridium jozwi (fragment)

N;Alternate names: endo-1,4-beta-glucanase  
C:Species: Clostridium jostii  
C>Date: 22-Jan-1993 #sequence\_revision 22-Jan-1993 #text\_change 22-Nov-1996  
C/Accession: A60156  
R:Fujino, T.; Sasaki, T.; Ohmura, K.; Shimizu, S.  
Appl. Environ. Microbiol. 56, 1175-1178, 1990  
A/Title: Purification and properties of an endo-1,4-beta-glucanase translated from a ClC  
A/Reference number: A60156; MUID:90253158; PMID:2187404  
A/Accession: A60156  
A/Molecule type: protein  
A/Residues: 1-15 <FU>  
C/Function:  
A/Description: hydrolysis of 1,4-beta-D-glucosidic linkages in beta-D-glucans such as ce  
C/Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 25.3%; Score 25; DB 2; Length 15;  
Best Local Similarity 46.2%; Pred. No. 1e+03;  
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 KDSLQMTDQYK 13  
DB 3 EDSHLLTNQAKK 15

RESULT 4  
S57584  
T cell receptor V-D-J junctional alpha chain region - human (fragment)  
C/Species: Homo sapiens (man)  
C/Date: 19-Oct-1995 #sequence\_revision 17-Nov-1995 #text\_change 05-Nov-1999  
C/Accession: S57584  
R:Burrows, S.R.; Silins, S.L.; Moss, D.J.; Khanna, R.; Mleko, I.S.; Argset, V.P.  
submitted to the EMBL Data Library, June 1995  
A/Description: T cell receptor repertoire for a viral epitope in humans is diversified b  
A/Reference number: S57494  
A/Accession: S57584  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-15 <BUR>  
A/Cross-references: EMBL:Z49956; NID:9887466; PIDN:CA90227.1; PID:9887467  
C/Keywords: T-cell receptor

Query Match 23.2%; Score 23; DB 2; Length 15;  
Best Local Similarity 66.7%; Pred. No. 2.1e+03;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 13 KESDDK 18  
DB 6 RERDDK 11

RESULT 5  
G35141  
T-cell receptor delta chain V region (105-7) - mouse (fragment)  
C/Species: Mus musculus (house mouse)  
C/Date: 31-Aug-1990 #sequence\_revision 31-Aug-1990 #text\_change 30-May-1997  
C/Accession: G35141  
R:Sim, G.K.; Auguetin, A.  
Cell 61, 397-405, 1990  
A/Title: Dominantly inherited expression of BID, an invariant undiversified T cell recep  
A/Reference number: A35141; MUID:90242386; PMID:2110506  
A/Accession: G35141  
A/Status: preliminary; not compared with conceptual translation  
A/Molecule type: mRNA  
A/Residues: 1-15 <SIM>  
C/Keywords: T-cell receptor

Query Match 23.2%; Score 23; DB 2; Length 15;  
Best Local Similarity 37.5%; Pred. No. 2.1e+03;  
Matches 3; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 11 YNKESDDK 18  
DB 11 YNKESDDK 18

DB 7 YRRDTSK 14

RESULT 6  
S52125  
gamma2-gliadin P25-27 - poulard wheat  
C/Species: Triticum turgidum (poulard wheat)  
C/Date: 14-Jul-1995 #sequence\_revision 03-Nov-1995 #text\_change 16-Feb-1997  
C/Accession: S52125  
R:Roche, A.; Soriano, F.; Molina, E.; Gonzalez-Linas, G.; Mender, E.  
Biochim. Biophys. Acta 1247, 143-148, 1995  
A/Title: Characterization of distinct alpha- and gamma-type gliadins and low molecular we  
A/Reference number: S52124; MUID:95178530; PMID:7873584  
A/Accession: S52125  
A/Status: preliminary  
A/Molecule type: protein  
A/Residues: 1-18 <ROC>  
C/Superfamily: gliadin

Query Match 23.2%; Score 23; DB 2; Length 18;  
Best Local Similarity 42.9%; Pred. No. 2.6e+03;  
Matches 3; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 LQWITDQ 10  
DB 10 VQWLOQ 16

RESULT 7  
PQ0175  
stylar glycoprotein 7 - Persian tobacco (fragment)  
N;Alternate names: glycoprotein S7  
C/Species: Nicotiana glauca (Persian tobacco)  
C/Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 13-Mar-1997  
C/Accession: PQ0175  
R:Jahnen, W.; Batterham, M.P.; Clarke, A.E.; Moritz, R.L.; Simpson, R.  
Plant Cell 1, 493-499, 1989  
A/Title: Identification, isolation, and N-terminal sequencing of stylar glycoproteins asac  
A/Reference number: PQ0173; MUID:92404717; PMID:2555548  
A/Accession: PQ0175  
A/Molecule type: protein  
A/Residues: 1-15 <JAH>  
A/Experimental source: style  
A/Comment: This protein is involved in self-incompatibility of flowering plants.  
C/Superfamily: Enterobacter ribonuclease  
C/Keywords: glycoprotein

Query Match 22.2%; Score 22; DB 2; Length 15;  
Best Local Similarity 80.0%; Pred. No. 2.9e+03;  
Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 LQWIT 8  
DB 9 LQWPT 13

RESULT 8  
C85956  
hypothetical protein Z4327 [imported] - Escherichia coli (strain O157:H7, substrain EDL93  
C/Species: Escherichia coli  
C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C/Accession: C85956  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,  
Miller, U.; Grobeck, B.D.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A/Reference number: A85480; MUID:21074935; PMID:11206551  
A/Accession: C85956  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-17 <STO>  
A/Cross-references: GB:AB005174; NID:912517535; PIDN:AAG58111.1; GSPDB:GN00145; UWGP:Z432  
A/Experimental source: strain O157:H7, substrain EDL933

C:Genetics:  
A:Gene: Z4327

Query Match 22.2%; Score 22; DB 2; Length 17;  
Best Local Similarity 42.9%; Pred. No. 3.4e+03;  
Matches 3; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 11 YNKESDD 17  
: : : : :  
Db 3 FDKSTDD 9

## RESULT 9

A39040  
C:Accession: A39040  
C:Species: Canis lupus familiaris (dog)  
C:Date: 23-Aug-1991 #sequence\_revision 23-Aug-1991 #text\_change 23-Feb-1997  
R:Calá, S.E.; Jones, L.R.  
J. Biol. Chem. 266, 391-398, 1991  
A:Title: Phosphorylation of cardiac and skeletal muscle calsequestrin isoforms by casein  
A:Reference number: A39040; PMID:91093153; PMID:1985907  
A:Accession: A39040  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 1-18 <CAL>  
C:Keywords: cardiac muscle; heart; phosphoprotein

Query Match 22.2%; Score 22; DB 2; Length 18;  
Best Local Similarity 66.7%; Pred. No. 3.6e+03;  
Matches 4; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 12 NKESDD 17  
: : : : :  
Db 12 NTEDD 17

## RESULT 10

PX0081  
C:Accession: PX0081  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 22-Apr-1995 #sequence\_revision 26-May-1995 #text\_change 07-May-1999  
R:Ohkubo, I.; Huang, K.; Ochiai, Y.; Takagaki, M.; Kan, K.  
J. Biochem. 116, 1182-1186, 1994  
A:Title: Dipeptidyl peptidase IV from porcine seminal plasma: Purification, characterization  
A:Reference number: PX0081; PMID:95204396; PMID:7896751  
A:Accession: PX0081  
A:Molecule type: protein  
A:Residues: 1-18 <OHK>  
A:Experimental source: seminal plasma  
A:Comment: This enzyme is a serine peptidase that preferentially liberates Xaa-proline  
C:Keywords: dipeptidylpeptidase hydrolase

Query Match 22.2%; Score 22; DB 2; Length 18;  
Best Local Similarity 66.7%; Pred. No. 3.6e+03;  
Matches 4; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 12 NKESDD 17  
: : : : :  
Db 1 NKGTDD 6

## RESULT 11

B60763  
C:Accession: B60763  
C:Species: Bacillus circulans  
C:Date: 14-May-1993 #sequence\_revision 14-May-1993 #text\_change 07-May-1999  
R:Flake, M.J.; Tobey-Fincher, K.L.; Fuchs, R.L.  
J. Gen. Microbiol. 136, 2377-2383, 1990  
A:Title: Cloning of two genes from Bacillus circulans WL-12 which encode 1,3-beta-glucan  
A:Reference number: A60763; PMID:91178514; PMID:2127800

A:Accession: B60763  
A:Molecule type: protein  
A:Residues: 1-15 <FIS>  
C:Comment: This bacillus produces up to six different 1,3-beta-glucanases for growth on C:Keywords: glycosidase; hydrolase

Query Match 21.7%; Score 21.5; DB 2; Length 15;  
Best Local Similarity 40.0%; Pred. No. 3.5e+03;  
Matches 4; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 3 SLOWITDYN 12  
: : : : :  
Db 5 NLWV-SDEFN 13

## RESULT 12

RHLMGs  
C:Accession: A01412  
C:Species: Petromyzon marinus (sea lamprey)  
C:Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 18-Mar-1997  
R:Sherwood, N.M.; Sower, S.A.; Marshak, D.R.; Frazer, B.A.; Brownstein, M.J.  
J. Biol. Chem. 261, 4812-4819, 1986  
A:Title: Primary structure of gonadotropin-releasing hormone from lamprey brain.  
A:Reference number: A01412; PMID:86168192; PMID:3514603  
A:Accession: A01412  
A:Molecule type: protein  
A:Residues: 1-10 <SHR>  
C:Comment: This hormone was isolated from the brain.  
C:Superfamily: gonadoliberin  
C:Keywords: amidated carboxyl end, hormone, pyroglutamic acid  
F:1/Modified site: pyroglutamate carboxylic acid (Gln) #status experimental  
F:10/Modified site: amidated carboxyl end (Gly) #status experimental

Query Match 21.2%; Score 21; DB 1; Length 10;  
Best Local Similarity 75.0%; Pred. No. 2.7e+03;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 SLOW 6  
: : : : :  
Db 4 SLEW 7

## RESULT 13

PC1215  
C:Accession: PC1215  
C:Species: Echinosoccus granulatus  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 30-Sep-1993  
R:Oliver, G.; Viipo, M.; Mailhos, A.; Martinez, C.; Sosa-Pineda, B.; Fiehlitz, W.; Ehrlich  
Gene 121, 337-342, 1992  
A:Title: Homeoboxes in flatworms.  
A:Reference number: JCI386; PMID:93077050; PMID:1359988  
A:Accession: PC1215  
A:Molecule type: DNA  
A:Residues: 1-14 <OLI>  
A:Cross-references: EMBL:X66821  
C:Keywords: homeobox

Query Match 21.2%; Score 21; DB 2; Length 14;  
Best Local Similarity 33.3%; Pred. No. 3.9e+03;  
Matches 3; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 6 WITDOYKE 14  
: : : : :  
Db 2 WFOYRSKE 10

## RESULT 14

B83636  
C:Accession: B83636  
C:Species: Bacillus halodurans  
C:Date: 14-May-1993 #sequence\_revision 14-May-1993 #text\_change 07-May-1999  
R:Flake, M.J.; Tobey-Fincher, K.L.; Fuchs, R.L.  
J. Gen. Microbiol. 136, 2377-2383, 1990  
A:Title: Cloning of two genes from Bacillus halodurans (strain C-125)  
A:Reference number: A60763; PMID:91178514; PMID:2127800

C:/Date: 01-Dec-2000 #sequence\_reviseion 01-Dec-2000 #text\_change 15-Jun-2001  
 C:/Accession: B83836  
 R/Takam, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai  
 Nucleic Acids Res. 28, 4317-4331, 2000  
 A:/Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
 A:/Reference number: A83650; MUID:20512582; PMID:11058132  
 A:/Accession: B83836  
 A:/Status: preliminary  
 A:/Molecule type: DNA  
 A:/Residues: 1-14 <STO>  
 A:/Cross-references: GB:AP001512; GB:BA000004; NID:q10174030; PIDN:BA05209.1; GSPDB:GNOC  
 A:/Experimental source: strain C-125  
 C:/Genetics:  
 A:/Gene: BH1490

Query Match 21.2%; Score 21; DB 2; Length 14;  
 Best Local Similarity 42.9%; Pred. No. 3.9e+03;  
 Matches 3; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 11 YNKESDD 17  
 |||  
 Db 4 YQKEER 10

## RESULT 15

S21238  
 hydrogensulfite reductase (EC 1.8.9.9.3) beta chain - *Desulfovibrio vulgaris* (fragment)  
 N/Alternate names: bisulfite reductase; desulfotuscidin; desulfovibridin; desulfovibridin;  
 C/Species: *Desulfovibrio vulgaris*  
 C:/Date: 19-Mar-1997 #sequence\_reviseion 11-Jun-1999 #text\_change 11-Jun-1999  
 C:/Accession: S21238  
 R/Pierik, A.J.; Duyvis, M.G.; van Helvoort, J.M.L.M.; Wolbert, R.B.G.; Hagen, W.R.  
 Eur. J. Biochem. 205, 111-115, 1992  
 A:/Title: The third subunit of desulfovibridin-type dissimilatory sulfite reductases.  
 A:/Reference number: S21197; MUID:92209491; PMID:1555572  
 A:/Accession: S21238  
 A:/Molecule type: protein  
 A:/Residues: 1-15 <PIE>  
 A:/Experimental source: strain Hildenborough  
 C:/Genetics:  
 A:/Gene: dsvB  
 C:/Complex: heterohexamer; two alpha, two beta and two gamma chains  
 C:/Function:  
 A:/Description: catalyzes the six-electron reduction of sulfite to sulfide  
 A:/Pathway: the terminal oxidase in the sulfate-reduction pathway  
 C:/Keywords: heterohexamer; oxidoreductase

Query Match 21.2%; Score 21; DB 2; Length 15;  
 Best Local Similarity 44.4%; Pred. No. 4.2e+03;  
 Matches 4; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 6 WITDYNKE 14  
 |||  
 Db 2 FISSGYNPE 10

Search completed: January 21, 2004, 12:20:17  
 Job time : 22 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: January 21, 2004, 12:08:45 ; Search time 6.3871 Seconds

(without alignments)  
576.265 Million cell updates/sec

Title: US-09-941-314-7

Perfect score: 99  
Sequence: 1 KDSIQWITTDQYNKESDDK 18Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 762491 seqs, 204481190 residues

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99	100.0	18	10	US-09-941-314-7 Sequence 7, Appli
2	99	100.0	35	10	US-09-941-314-8 Sequence 8, Appli
3	99	100.0	36	10	US-09-941-314-6 Sequence 6, Appli
4	99	100.0	50	9	US-09-864-761-34822 Sequence 34822, A
5	99	100.0	50	9	US-09-864-761-48936 Sequence 48936, A
6	99	100.0	115	10	US-09-941-314-3 Sequence 3, Appli
7	99	100.0	117	10	US-09-941-314-4 Sequence 4, Appli
8	99	100.0	137	10	US-09-941-314-2 Sequence 2, Appli
9	55	55.6	27	10	US-09-941-314-5 Sequence 9, Appli
10	55	55.6	49	10	US-09-941-314-13 Sequence 13, Appli
11	49	49.5	80	10	US-09-941-314-15 Sequence 15, Appli
12	46	46.5	481	9	US-09-815-242-5584 Sequence 5584, Ap
13	46	46.5	487	9	US-09-815-242-12456 Sequence 12456, A
14	44	44.4	418	12	US-10-359-493-10941 Sequence 10941, A
15	44	44.4	481	9	US-09-815-242-13425 Sequence 13425, A

16	44	44.4	482	9	US-09-815-242-4953 Sequence 4953, Ap
17	44	44.4	491	9	US-09-815-242-10940 Sequence 10940, A
18	44	44.4	655	14	US-10-117-641-36 Sequence 36, Appli
19	44	44.4	655	15	US-10-235-113-36 Sequence 36, Appli
20	43	43.4	436	12	US-10-369-493-19413 Sequence 19413, A
21	43	43.4	485	12	US-10-369-493-17102 Sequence 2, Appli
22	43	43.4	2285	10	US-09-932-1834-2 Sequence 278, App
23	42.5	42.9	715	12	US-10-238-075-278 Sequence 3, Appli
24	42.5	42.9	362	12	US-10-316-067-3 Sequence 3, Appli
25	42	42.4	362	15	US-10-235-056-3 Sequence 5, Appli
26	42	42.4	363	12	US-10-316-067-5 Sequence 5, Appli
27	42	42.4	363	15	US-10-235-056-5 Sequence 5, Appli
28	42	42.4	448	12	US-10-340-779A-4 Sequence 749A, App
29	42	42.4	466	12	US-10-032-585-7449 Sequence 586, App
30	42	42.4	976	12	US-10-117-937-596 Sequence 596, App
31	41	41.4	275	12	US-09-882-227-420 Sequence 420, App
32	41	41.4	315	10	US-09-738-626-5497 Sequence 5497, App
33	41	41.4	805	12	US-10-369-493-6498 Sequence 6498, App
34	41	41.4	805	12	US-10-369-493-6499 Sequence 6499, App
35	41	41.4	816	9	US-09-815-242-13947 Sequence 13947, A
36	41	41.4	880	12	US-10-369-493-8129 Sequence 8129, App
37	40.5	40.9	173	9	US-09-205-658-214 Sequence 214, App
38	40.5	40.9	173	12	US-09-963-693-214 Sequence 214, App
39	40	40.4	134	11	US-09-969-730-114 Sequence 114, App
40	40	40.4	134	11	US-09-969-730-199 Sequence 199, App
41	40	40.4	135	11	US-09-774-639-114 Sequence 114, App
42	40	40.4	165	15	US-10-106-698-5512 Sequence 5512, App
43	40	40.4	234	12	US-10-369-493-22270 Sequence 22270, A
44	40	40.4	235	12	US-10-383-834-5 Sequence 5, Appli
45	40	40.4	370	15	US-10-156-761-13298 Sequence 13298, A

## ALIGNMENTS

RESULT 1  
US-09-941-314-7  
Sequence 7, Application US/09941314  
Patent No. US20020142396A1  
GENERAL INFORMATION:  
APPLICANT: ZymoGenetics, Inc.  
TITLE OF INVENTION: Mammalian Cystatin-B and Its Use to  
FILE REFERENCE: 00-81PC  
CURRENT FILING DATE: 2001-08-29  
PRIOR APPLICATION NUMBER: 60/230,230  
PRIOR FILING DATE: 2001-09-01  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 7  
LENGTH: 18  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-941-314-7

Query Match 100.0%; Score 99; DB 10; Length 18;  
Best Local Similarity 100.0%; Pred. No. 6.7e-08;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 KDSIQWITTDQYNKESDDK 18  
DB 1 KDSIQWITTDQYNKESDDK 18

RESULT 2  
US-09-941-314-8  
Sequence 8, Application US/09941314  
Patent No. US20020142396A1  
GENERAL INFORMATION:  
APPLICANT: ZymoGenetics, Inc.  
TITLE OF INVENTION: Mammalian Cystatin-B and Its Use to  
FILE REFERENCE: 00-81PC  
CURRENT FILING DATE: 2001-08-29  
PRIOR APPLICATION NUMBER: 60/230,230  
PRIOR FILING DATE: 2001-09-01  
NUMBER OF SEQ ID NOS: 19  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 7  
LENGTH: 18  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-941-314-7

```

7 PRIOR FILING DATE: 2000-10-04
7 PRIOR APPLICATION NUMBER: US 60/236,359
7 PRIOR FILING DATE: 2000-09-27
7 PRIOR APPLICATION NUMBER: PCT/US01/00666
7 PRIOR FILING DATE: 2001-01-30
7 PRIOR APPLICATION NUMBER: PCT/US01/00667
7 PRIOR FILING DATE: 2001-01-30
7 PRIOR APPLICATION NUMBER: PCT/US01/00664
7 PRIOR FILING DATE: 2001-01-30
7 PRIOR APPLICATION NUMBER: PCT/US01/00665
7 PRIOR FILING DATE: 2001-01-30
7 PRIOR APPLICATION NUMBER: PCT/US01/00665
7 PRIOR FILING DATE: 2001-01-30
7 PRIOR APPLICATION NUMBER: PCT/US01/00668
7 PRIOR FILING DATE: 2001-01-30
7 PRIOR APPLICATION NUMBER: PCT/US01/00663
7 PRIOR FILING DATE: 2001-01-30
7 PRIOR APPLICATION NUMBER: PCT/US01/00662
7 PRIOR FILING DATE: 2001-01-30
7 PRIOR APPLICATION NUMBER: PCT/US01/00661
7 PRIOR FILING DATE: 2001-01-30
7 PRIOR APPLICATION NUMBER: PCT/US01/00670
7 PRIOR FILING DATE: 2001-01-30
7 PRIOR APPLICATION NUMBER: US 60/234,687
7 PRIOR FILING DATE: 2000-09-21
7 PRIOR APPLICATION NUMBER: US 09/608,408
7 PRIOR FILING DATE: 2000-06-30
7 PRIOR APPLICATION NUMBER: US 09/774,203
7 PRIOR FILING DATE: 2001-01-29
7 NUMBER OF SEQ ID NOS: 49117
7 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
7 SEQ ID NO 34822
7 LENGTH: 50
7 TYPE: PRT
7 ORGANISM: Homo sapiens
7 FEATURE:
7 OTHER INFORMATION: MAP TO AL109954.10
7 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
7 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
7 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.8
7 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.7
7 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
7 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
7 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
7 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
7 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
7 OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.2
7 OTHER INFORMATION: EST HUMAN HIT: A1200857.1, EVALUE 5.00e-23
7 OTHER INFORMATION: SWISSPROT HIT: O60676, EVALUE 1.00e-01
7 US-09-864-761-34822
7
7 Query Match 100.0%; Score 99; DB 9; Length 50;
7 Best Local Similarity 100.0%; Pred. No. 1.9e-07;
7 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
7
7 QY 1 KDSLOWITDYNKESDDK 18
7 |||||
7 |||||
7 |||||
7 |||||
7 |||||
7 DB 19 KDSLOWITDYNKESDDK 36
7
7 RESULT 5
7 US-09-864-761-48936
7 Sequence 48936, Application US/09864761
7 Patent No. US20020048763A1
7 GENERAL INFORMATION:
7 APPLICANT: Penn, Sharon G.
7 APPLICANT: Rank, David R.
7 APPLICANT: Hanzel, David K.
7 APPLICANT: Chen, Wensheng
7 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
7 TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
7 FILE REFERENCE: Aeomica-X-1
7 CURRENT APPLICATION NUMBER: US/09/864,761

```

```
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 48936
;; LENGTH: 50
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL096677.18
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.96
;; OTHER INFORMATION: EST HUMAN HIT: AI200857.1, EVALU8 5.00e-23
;; OTHER INFORMATION: SWISSPROT HIT: O60676, EVALU8 1.00e-01
US-09-864-761-48936

Query Match          100.0%; Score 99; DB 9; Length 50;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 KDSLOWITDOYNKESDDK 18
Db      19 KDSLOWITDOYNKESDDK 36
```

```
RESULT 6
US-09-941-314-3
; Sequence 3, Application US/09941314
; Patent No. US20020142396A1
; GENERAL INFORMATION:
; APPLICANT: ZymoGenetics, Inc.
; TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to
; TITLE OF INVENTION: Inhibit Cancer Procoagulant Protein
; FILE REFERENCE: 00-81PC
; CURRENT APPLICATION NUMBER: US/09/941,314
; CURRENT FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/230,230
; PRIOR FILING DATE: 2001-09-01
; NUMBER OF SEQ ID NOS: 19
```

```
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 3
;; LENGTH: 115
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-941-314-3
```

```
Query Match          100.0%; Score 99; DB 10; Length 115;
Best Local Similarity 100.0%; Pred. No. 4.3e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 KDSLOWITDOYNKESDDK 18
Db      22 KDSLOWITDOYNKESDDK 39
```

```
RESULT 7
US-09-941-314-4
; Sequence 4, Application US/09941314
; Patent No. US20020142396A1
; GENERAL INFORMATION:
; APPLICANT: ZymoGenetics, Inc.
; TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to
; TITLE OF INVENTION: Inhibit Cancer Procoagulant Protein
; FILE REFERENCE: 00-81PC
; CURRENT APPLICATION NUMBER: US/09/941,314
; CURRENT FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/230,230
; PRIOR FILING DATE: 2001-09-01
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-941-314-4
```

```
Query Match          100.0%; Score 99; DB 10; Length 117;
Best Local Similarity 100.0%; Pred. No. 4.4e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 KDSLOWITDOYNKESDDK 18
Db      24 KDSLOWITDOYNKESDDK 41
```

```
RESULT 8
US-09-941-314-2
; Sequence 2, Application US/09941314
; Patent No. US20020142396A1
; GENERAL INFORMATION:
; APPLICANT: ZymoGenetics, Inc.
; TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to
; TITLE OF INVENTION: Inhibit Cancer Procoagulant Protein
; FILE REFERENCE: 00-81PC
; CURRENT APPLICATION NUMBER: US/09/941,314
; CURRENT FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/230,230
; PRIOR FILING DATE: 2001-09-01
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 137
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-941-314-2
```

```
Query Match          100.0%; Score 99; DB 10; Length 137;
Best Local Similarity 100.0%; Pred. No. 5.2e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 KDSLOWITDOYNKESDDK 18
```

Db 44 KDSLOWITDQYNKESDDK 61

RESULT 9

US-09-941-314-9

Sequence 9, Application US/09941314

Patent No. US20020142396A1

GENERAL INFORMATION:

APPLICANT: ZymoGenetics, Inc.

TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to

FILE REFERENCE: 00-81PC

CURRENT APPLICATION NUMBER: US/09/941,314

CURRENT FILING DATE: 2001-08-29

PRIOR APPLICATION NUMBER: 60/230,230

PRIOR FILING DATE: 2001-09-01

NUMBER OF SEQ ID NOS: 19

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 9

LENGTH: 27

TYPE: PRT

ORGANISM: Homo sapiens

US-09-941-314-9

Query Match 55.6%; Score 55; DB 10; Length 27;

Best Local Similarity 100.0%; Pred. No. 0.26; Mismatches 0; Indels 0; Gaps 0;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DQYNKESDDK 18

Db 1 DQYNKESDDK 10

RESULT 10

US-09-941-314-13

Sequence 13, Application US/09941314

Patent No. US20020142396A1

GENERAL INFORMATION:

APPLICANT: ZymoGenetics, Inc.

TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to

FILE REFERENCE: 00-81PC

CURRENT APPLICATION NUMBER: US/09/941,314

CURRENT FILING DATE: 2001-08-29

PRIOR APPLICATION NUMBER: 60/230,230

PRIOR FILING DATE: 2001-09-01

NUMBER OF SEQ ID NOS: 19

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 13

LENGTH: 49

TYPE: PRT

ORGANISM: Homo sapiens

US-09-941-314-13

Query Match 55.6%; Score 55; DB 10; Length 49;

Best Local Similarity 100.0%; Pred. No. 0.47; Mismatches 0; Indels 0; Gaps 0;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 DQYNKESDDK 18

Db 1 DQYNKESDDK 10

RESULT 11

US-09-941-314-15

Sequence 15, Application US/09941314

Patent No. US20020142396A1

GENERAL INFORMATION:

APPLICANT: ZymoGenetics, Inc.

TITLE OF INVENTION: Mammalian Cystatin-8 and Its Use to

FILE REFERENCE: 00-81PC

CURRENT APPLICATION NUMBER: US/09/941,314

CURRENT FILING DATE: 2001-08-29

PRIOR APPLICATION NUMBER: 60/230,230

PRIOR FILING DATE: 2001-09-01

NUMBER OF SEQ ID NOS: 19

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 15

LENGTH: 80

TYPE: PRT

ORGANISM: Homo sapiens

US-09-941-314-15

Query Match 49.5%; Score 49; DB 10; Length 80;

Best Local Similarity 100.0%; Pred. No. 5.8; Mismatches 0; Indels 0; Gaps 0;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 QYNKESDDK 18

Db 1 QYNKESDDK 9

RESULT 12

US-09-815-242-5584

Sequence 5584, Application US/09815242

Patent No. US2002061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.

APPLICANT: Zyckind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21,078

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 5584

LENGTH: 481

TYPE: PRT

ORGANISM: Staphylococcus aureus

US-09-815-242-5584

Query Match 46.5%; Score 46; DB 9; Length 481;

Best Local Similarity 43.8%; Pred. No. 97; Mismatches 6; Indels 0; Gaps 0;

Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 KDSLOWITDQYNKESD 16

Db 318 KQKLAWVNNQYKQKD 333

RESULT 13

US-09-815-242-12456

Sequence 12456, Application US/09815242

Patent No. US2002061569A1

GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zykkind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12456  
LENGTH: 487  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-12456

Query Match 46.5%; Score 46; DB 9; Length 487;  
Best Local Similarity 43.8%; Pred. No. 98;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 1 KDSLOWITDOYKESD 16  
DB 321 KQKLAWNNQYMKQD 336

RESULT 14  
US-10-369-493-10941  
Sequence 10941, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianheng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 10941  
LENGTH: 418  
TYPE: PRT  
ORGANISM: Ferroplasma acidarmanus  
US-10-369-493-10941

Query Match 44.4%; Score 44; DB 12; Length 418;  
Best Local Similarity 57.1%; Pred. No. 1.6e+02;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 5 QWITDOYKESDK 18  
DB 33 QHITDYKESSEK 46

RESULT 15  
US-09-815-242-13425  
Sequence 13425, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zykkind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13425  
LENGTH: 481  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-815-242-13425

Query Match 44.4%; Score 44; DB 9; Length 481;  
Best Local Similarity 61.5%; Pred. No. 1.9e+02;  
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 4 LQWITDOYKESD 16  
DB 382 LQKITDAVNRDAD 394

Search completed: January 21, 2004, 12:17:40  
Job time: 7.3871 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 12:10:05 ; Search time 11 Seconds  
(without alignments)

76.953 Million cell updates/sec

Title: US-09-941-314-7

Sequence: 1 KOSLOWITTDQYRKESDDK 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 987

Minimum DB seq length: 0  
Maximum DB seq length: 18

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25.3	25.3	16	1	CAT9_FASHE
2	25.3	25.3	16	1	LE06_BIOGL
3	24.2	24.2	13	1	RPOC_MYCA
4	22.2	22.2	11	1	RANC_RANPI
5	22.2	22.2	17	1	RANR_RANPI
6	21.2	21.2	10	1	GONI_PETWA
7	21.2	21.2	10	1	SP34_DICMU
8	21.2	21.2	15	1	OBPA_MAMBR
9	19.2	19.2	10	1	AEGL_AGRAB
10	19.2	19.2	10	1	AL19_CARMA
11	19.2	19.2	10	1	UH05_RAT
12	19.2	19.2	12	1	NO40_SOYBN
13	19.2	19.2	16	1	LE05_BIOGL
14	19.2	19.2	16	1	TRYP_PELCA
15	19.2	19.2	17	1	LEW_CORGL
16	18.2	18.2	9	1	LEPC_STAUV
17	18.2	18.2	10	1	LSK2_LEDMA
18	18.2	18.2	10	1	MP2_MICOC
19	18.2	18.2	10	1	NO40_TOBAC
20	18.2	18.2	13	1	BI21_LITRU
21	18.2	18.2	13	1	BI22_LITRU
22	18.2	18.2	13	1	PIBB_RABIT
23	18.2	18.2	13	1	LEPA_PORGI
24	18.2	18.2	13	1	TEML_RANGI
25	18.2	18.2	15	1	BCDA_LYMDI
26	18.2	18.2	15	1	CR78_HORSE
27	18.2	18.2	15	1	PEKH_PHYPA
28	17.2	17.2	9	1	LITR_PHYRO
29	17.2	17.2	10	1	AMPN_HELM
30	17.2	17.2	10	1	PIBB_CERBI
31	17.2	17.2	11	1	CORZ_PERRM
32	17.2	17.2	11	1	PARI_CALVO
33	17.2	17.2	12	1	PARI_CALVO

34	17	17.2	13	1	BOML_PSECU	P42991 pseudophryn
35	17	17.2	13	1	UHA3_CANPA	P5535 canis fami1
36	17	17.2	14	1	GLGS_SPTOL	P55235 spinacia ol
37	17	17.2	15	1	LEC2_PSOOC	P22585 psophocarpu
38	17	17.2	15	1	PC20_BRANA	P81036 brassica na
39	17	17.2	18	1	PHPT_PSRSE	P25271 pseudalelia
40	17	17.2	18	1	RL24_PROVU	P20032 proteus vul
41	16	16.2	8	1	CCRN_MACEU	P30369 macropus eu
42	16	16.2	9	1	AL11_CARMA	P81814 carinus ma
43	16	16.2	9	1	PIBB_MACTU	P19345 macaca fusc
44	16	16.2	10	1	BPP8_BOTIN	P30426 botrops in
45	16	16.2	10	1	MAL8_KLEPN	Q05564 klebsiella

## ALIGNMENTS

RESULT 1	CAT9_FASHE	STANDARD;	PRT;	16 AA.
AC	P80533;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Putative cathepsin-like enzyme (EC 3.4.22.-) (Newly excysted juvenile protein 9) (Fragment).			
OS	Fasciola hepatica (Liver fluke).			
OC	Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Fasciolidae; Fasciola.			
OC	Echinostomida; Echinostomata; Fascioloidae; Fasciolidae; Fasciola.			
OX	NCBI_TaxID=6192;			
RN	[1]			
RP	SEQUENCE			
RX	MEDLINE=95366993; Pubmed=7639732;			
RA	Kalcevic J., Ashman K., Meusen B.;			
RT	"Fasciola hepatica: rapid identification of newly excysted juvenile proteins."			
RL	Biochem. Biophys. Res. Commun. 213:169-174(1995).			
CC	-1- DEVELOPMENTAL STAGE: EXPRESSED AT THE NEWLY EXCYSTED JUVENILE STAGE.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.			
DR	MEROPS; C01.033; -			
DR	InterPro; IPR000169; SHPOT acsite.			
DR	PROSITE; PS00139; THIOL_PROTEASE_CYS; PARTIAL.			
DR	PROSITE; PS00639; THIOL_PROTEASE_HIS; PARTIAL.			
DR	PROSITE; PS00640; THIOL_PROTEASE_ASN; PARTIAL.			
KW	Hydrolase; Thiol protease.			
FT	NON TER			
SQ	SEQUENCE 16 AA; 1966 MW; 0B7B18FDB1FA541B CRC64;			
QY	2 DSILOW 6	25.34; Score 25; DB 1; Length 16;		
DB	4 DSILOW 8	Best Local Similarity 80.0%; Pred. No. 5.2e+02;		
		Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
RESULT 2	LE06_BIOGL	STANDARD;	PRT;	16 AA.
ID	P80745;			
AC	P80745;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Hemolymph 65 kDa lectin BG06 (Fragment).			
GN	BG06.			
OS	Blomphalaria glabrata (Bloodfluke planorb).			
OC	Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;			
OC	Lymnaeidae; Planorbidae; Blomphalaria.			
OX	NCBI_TaxID=6526;			
RN	[1]			
RP	SEQUENCE.			

```

RC STRAIN=M-LINE; TISSUE=Hemolymph;
RX MEDLINE=97385165; PubMed=9238039;
RA Adema C.M., Hertel L.A., Miller R.D., Loker E.S.;
RT "A family of fibrinogen-related proteins that precipitates parasite-
derived molecules is produced by an invertebrate after infection.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:8691-8696(1997).
CC -1- FUNCTION: BINDS AND PRECIPITATES ANTIGENS OF THE PARASITE
CC ECHINOSTOMA PARASENSI.
CC -1- SUBCELLULAR LOCATION: SECRETED IN THE HEMOLYMPH.
CC -1- INDUCTION: By infection.
KW Lectin.
FT NON TER 1 1
FT NON TER 16 16
SQ SEQUENCE 16 AA; 1964 MW; A1665754589EF82C CRC64;

Query Match 25.3%; Score 25; DB 1; Length 16;
Best Local Similarity 37.5%; Pred. No. 5.2e+02;
Matches 3; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 11 YNKESDDK 18
DB 5 FDKNDQDQ 12

RESULT 3
RPOC MYCGA STANDARD; PRT; 13 AA.
AC P47716
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6) (Transcriptase
DE beta' chain) (RNA polymerase beta' subunit) (Fragment).
GN RPOC.
OS Mycoplasma gallisepticum.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2096;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AS5969Var.B;
RA Skamov A.V., Feoktistova E.S., Gol'dman M.A., Feoktistova E.S.,
RA Bibilashvili R.S.;
RL Submitted (XXX-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA) (N).
CC -1- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC BETA' CHAIN.
CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L38402; AAB40952.1; -
KW Transferase; DNA-directed RNA polymerase; Transcription.
FT NON TER 13 13
FT NON TER 13 13
SQ SEQUENCE 13 AA; 1630 MW; 4BEC27C7480D4333 CRC64;

Query Match 24.2%; Score 24; DB 1; Length 13;
Best Local Similarity 40.0%; Pred. No. 5.9e+02;
Matches 4; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 9 DQYNKESDDK 18
DB 2 DKLNNKNNK 11

```

```

RESULT 4
RANC RANPI STANDARD; PRT; 11 AA.
AC P08951
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ranatensin-C. (Northern leopard frog).
OS Rana pipiens (Northern leopard frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Rana.
OX NCBI_TaxID=8404;
RN [1]
RP SEQUENCE.
RC TISSUE=Skin secretion; PubMed=6141890;
RX MEDLINE=84131098; PubMed=6141890;
RA Nakajima T.;
RL Unpublished results, cited by:
RL Erpapner V., Erpapner G.F., Mazzanti G., Eudean R.;
RL Comp. Biochem. Physiol. 77C:99-108(1984).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Skin.
CC -1- SIMILARITY: BELONGS TO THE BOMBESIN/NEUROMEDIN B/RANATENSIN
CC FAMILY.
CC InterPro; IPR000874; Bombesin.
DR Pfam; PF02044; Bombesin; 1.
DR PROSITE; PS00257; BOMBESIN; 1.
KW Amphibian defense peptide; Bombesin family; Amidation.
FT MOD_RES 11 11
FT MOD_RES 11 11
SQ SEQUENCE 11 AA; 1304 MW; D6C9885A61ADC366 CRC64;

Query Match 22.2%; Score 22; DB 1; Length 11;
Best Local Similarity 42.9%; Pred. No. 1e+03;
Matches 3; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 QWITDQY 11
DB 4 QWATGHR 10

RESULT 5
RANR RANRU STANDARD; PRT; 17 AA.
AC P08952
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ranatensin-R.
OS Rana rugosa (Wrinkled frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidae; Rana.
OX NCBI_TaxID=8410;
RN [1]
RP SEQUENCE.
RC TISSUE=Skin secretion; PubMed=6141890;
RX MEDLINE=84131098; PubMed=6141890;
RA Nakajima T.;
RL Unpublished results, cited by:
RL Erpapner V., Erpapner G.F., Mazzanti G., Eudean R.;
RL Comp. Biochem. Physiol. 77C:99-108(1984).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Skin.
CC -1- SIMILARITY: BELONGS TO THE BOMBESIN/NEUROMEDIN B/RANATENSIN
CC FAMILY.
CC InterPro; IPR000874; Bombesin.
DR Pfam; PF02044; Bombesin; 1.
DR PROSITE; PS00257; BOMBESIN; 1.
KW Amphibian defense peptide; Bombesin family; Amidation.
FT MOD_RES 17 17
FT MOD_RES 17 17
SQ SEQUENCE 17 AA; 2053 MW; 3A876B35A581863E CRC64;

```

Query Match 22.2%; Score 21; DB 1; Length 17;  
 Best Local Similarity 42.9%; Pred. No. 1.6e+03;  
 Matches 3; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 5 QMITDQY 11  
 |||:  
 Db 10 QMATHF 16

## RESULT 6

GON1\_PETMA STANDARD; PRT; 10 AA.

AC P04378; 20-MAR-1987 (Rel. 04, Created)  
 DT 20-MAR-1987 (Rel. 04, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Gonadolibertin I (Gonadotropin-releasing hormone I) (GNRH-I)  
 OS Petromyzon marinus (Sea lamprey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;  
 OC Petromyzontiformes; Petromyzontidae; Petromyzon.  
 OX NCBI\_TaxID=7757;  
 RN [1]  
 RP SEQUENCE.

RC TISSUE=Brain;  
 RX MEDLINE=66168192; PubMed=3514603;  
 RA Sherwood N.M., Sower S.A., Marshak D.R., Fraser B.A., Brownstein M.J.;  
 RT "Primary structure of gonadotropin-releasing hormone from lamprey brain."  
 RT J. Biol. Chem. 261:4812-4819(1986).  
 CC -1- FUNCTION: Stimulates the secretion of gonadotropins; it stimulates the secretion of both luteinizing and follicle-stimulating hormones.

CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: Belongs to the GNRH family.  
 DR InterPro: IPR002012; GNRH.  
 DR Pfam: PF00446; GNRH; 1.

DR PROSITE: PS00473; GNRH; 1.  
 KM Hormone; Amidation; Hypothalamus; Pyroglutamate carboxylic acid.  
 FT MOD\_RES 1 1 PYROGLUTAMATE CARBOXYLIC ACID.  
 FT MOD\_RES 10 10 AMIDATION.  
 SQ SEQUENCE 10 AA; 1244 MW; 1E4B36237B1735AB CRC64;

Query Match 21.2%; Score 21; DB 1; Length 10;  
 Best Local Similarity 75.0%; Pred. No. 1.3e+03;  
 Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 3 SLOW 6  
 |||:  
 Db 4 SLEW 7

## RESULT 7

SP34\_DICMU STANDARD; PRT; 10 AA.

AC P81545; 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Surface protein P34 (Fragment).  
 GN P34.  
 OS Dictyostelium mucoroides (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostelidae; Dictyostelium.  
 OX NCBI\_TaxID=31287;  
 RN [1]  
 RP SEQUENCE.

RC STRAIN=DM-7;  
 RA Schreiner S.J.;  
 RT "Characterization of a surface protein in macrocytes of Dictyostelium RT mucoroides."  
 RT Submitted (NOV-1998) to the SWISS-PROT data bank.  
 CC -1- FUNCTION: THIS PROTEIN IS PRESENT IN THE MACROCYST PRIMARY WALL

CC WHICH IS PRODUCED BY AMOEBAE DURING THE ONSET OF SEXUAL  
 CC REPRODUCTION.  
 CC -1- SUBCELLULAR LOCATION: MACROCYST PRIMARY WALL.  
 CC Cell wall.  
 KM NON TER 10 10  
 FT SEQUENCE 10 AA; 1190 MW; 1B6A707AA345B50 CRC64;

Query Match 21.2%; Score 21; DB 1; Length 10;  
 Best Local Similarity 50.0%; Pred. No. 1.3e+03;  
 Matches 3; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 11 YNKESD 16  
 |||:  
 Db 4 YNKDGN 9

## RESULT 8

OBPA\_MAMBR STANDARD; PRT; 15 AA.

AC P81285; 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Antennal odorant-binding protein (AOBP) (Fragment).  
 OS Mamestra brassicae.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Noctuoidea;  
 OC Noctuidae; Hadeninae; Mamestra.  
 OX NCBI\_TaxID=55057;  
 RN [1]  
 RP SEQUENCE.

RC TISSUE=Antenna;  
 RX MEDLINE=99097262; PubMed=9878563;  
 RA Bobbot J., Sobrio F., Lucas P., Nagnan-le Meunier P.;  
 RT "Functional characterization of a new class of odorant-binding proteins in the moth Mamestra brassicae."  
 RT Biochem. Biophys. Res. Commun. 253:489-494(1998).  
 CC -1- TISSUE SPECIFICITY: ANTENNA.  
 DR InterPro: IPR005055; A10\_OS-D.  
 DR Pfam: PF03392; OS-D; 1.

KM Olfaction; Transport.  
 FT NON TER 15  
 SQ SEQUENCE 15 AA; 1875 MW; 1EB9B36E5A74BAA CRC64;

Query Match 21.2%; Score 21; DB 1; Length 15;  
 Best Local Similarity 36.4%; Pred. No. 2e+03;  
 Matches 4; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 8 TDQYNKESDK 18  
 |||:  
 Db 5 TDKYDNIINLDE 15

## RESULT 9

AEGL\_AGRAB STANDARD; PRT; 10 AA.

AC P83465; 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Aegerolysin (Fragment).  
 OS Aegerolysin aegerita (Black poplar mushroom).  
 OC Eukaryota; Fungi; Basidiomycota; Hyenomycetes; Homobasidiomycetes;  
 OC Agaricales; Boletaceae; Agrocybe.  
 OX NCBI\_TaxID=5400;  
 RN [1]  
 RP SEQUENCE, FUNCTION, AND SUBUNIT.

RC STRAIN=PAV2 98; TISSUE=fruiting body;  
 RX MEDLINE=22015236; PubMed=12020804;  
 RA Berne S., Krizaj I., Pohlmann F., Turk T., Macek P., Sepcic K.;  
 RT "Pleurotus and Agrocybe hemolysins, new proteins hypothetically involved in fungal fruiting."  
 RT Biochim. Biophys. Acta 1570:153-159(2002).

CC -1- FUNCTION: Has hemolytic activity against bovine erythrocytes at  
 CC nanomolar concentrations. May play an important role in the  
 CC initial phase of fungal fruiting.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SIMILARITY: BELONGS TO THE AGEROLYSIN FAMILY.  
 KM Hemolysis.  
 FT DOMAIN 7 10 POLY-ILE.  
 FT NON\_TER 10 10  
 SQ SEQUENCE 10 AA, 1189 MW, DB7D55542D366DD CRC64;  
 Query Match 19.2%; Score 19; DB 1; Length 10;  
 Best Local Similarity 66.7%; Pred. No. 2.6e+03;  
 Matches 2; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 5 QW 7  
 Db 4 QW 6

RESULT 10  
 AL19\_CARMA STANDARD; PRT; 10 AA.  
 ID AL19\_CARMA  
 AC P81822;  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DE Carcinus maenas (Common shore crab) (Green crab).  
 OS Carcinus maenas (Common shore crab) (Green crab).  
 OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;  
 OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;  
 OC Eubrachyura; Portunidea; Portunidae; Carcinus.  
 OX NCBI\_TaxID=6759;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Cerebral ganglion, and Thoracic ganglion;  
 RX MEDLINE=98121193; PubMed=9461295;  
 RA Dye H., Johnsen A.H., Maestro J.-L., Scott A.G., Jaros P.F.,  
 RA Thorpe A.;  
 RT "Isolation and identification of multiple neuropeptides of the  
 RT allatostatin superfamily in the shore crab Carcinus maenas.";  
 RL Eur. J. Biochem. 250:727-734(1997).  
 CC -1- FUNCTION: MAY ACT AS A NEUROTRANSMITTER OR NEUROMODULATOR.  
 CC -1- SIMILARITY: BELONGS TO THE ALLATOSTATIN FAMILY.  
 KM Neuropeptide; Amidation; Multigene family.  
 FT MOD\_RES 10 10 AMIDATION (POTENTIAL).  
 SQ SEQUENCE 10 AA, 1101 MW, 96687CDSAB569AB1 CRC64;  
 Query Match 19.2%; Score 19; DB 1; Length 10;  
 Best Local Similarity 60.0%; Pred. No. 2.6e+03;  
 Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 8 TDQYN 12  
 Db 3 TDQYS 7

RESULT 11  
 UH05\_RAT STANDARD; PRT; 10 AA.  
 ID UH05\_RAT  
 AC P56573;  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DE Unknown protein from 2D-page of heart tissue (Spot P5) (Fragment).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE.  
 RC STEAIN=Wistar; TISSUE=Heart;  
 RA Li X.-P., Pleissner K.-P., Scheler C., Reglitz-Zagrosek V., Salikov J.,  
 RA Jungblut P.R.;

RL Submitted (SEP-1998) to the SWISS-PROT data bank.  
 CC -1- MISCELLANEOUS: ON THE 2D-GEL, THE DETERMINED PI OF THIS UNKNOWN  
 CC PROTEIN IS: 8.3, ITS MW IS: 30 KDa.  
 FT UNSURE 9 9 OR P.  
 FT NON\_TER 10 10  
 SQ SEQUENCE 10 AA, 1269 MW, 00CBA4B46C5BAB CRC64;  
 Query Match 19.2%; Score 19; DB 1; Length 10;  
 Best Local Similarity 42.9%; Pred. No. 2.6e+03;  
 Matches 3; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 10 QYNKESD 16  
 Db 2 QYDSQYD 8

RESULT 12  
 NO40\_SOYBN STANDARD; PRT; 12 AA.  
 ID NO40\_SOYBN  
 AC P55960;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Early nodulin 40.  
 GN ENOD40.  
 OS Glycine max (Soybean).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
 OX NCBI\_TaxID=3847;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Williams;  
 RX MEDLINE=94035161; PubMed=8220464;  
 RA Yang W.C., Kalinakis P., Hendrix P., Smolders A., de Vries F.,  
 RA Spee J., van Kammen A., Bisseling T., Franssen H.;  
 RT "Characterization of GmENOD40, a gene showing novel patterns of cell-  
 RT specific expression during soybean nodule development.";  
 RL Plant J. 3:573-585(1993)  
 CC -1- FUNCTION: MODULATES THE ACTION OF AUXIN, AND MAY FUNCTION AS PLANT  
 CC GROWTH REGULATOR THAT ALTERS PHYTOHORMONE RESPONSES (BY  
 CC SIMILARITY).  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE EARLY STAGES OF THE NODULE  
 CC DEVELOPMENT.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 CC EMBL; X69154; -; NOT\_ANNOTATED\_CDS.  
 DR Nodulation.  
 KM Nodulation.  
 SQ SEQUENCE 12 AA, 1391 MW, 3C6958AE78B1A733 CRC64;  
 Query Match 19.2%; Score 19; DB 1; Length 12;  
 Best Local Similarity 60.0%; Pred. No. 3.1e+03;  
 Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 4 LQWIT 8  
 Db 3 LQWIT 7

RESULT 13  
 LB05\_BIOGL STANDARD; PRT; 16 AA.  
 ID LB05\_BIOGL  
 AC P80744;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hemolymph 65 kDa lectin BG05 (Fragment).  
 GN BG05.  
 OS Blomphalaria glabrata (Biodidituke planorb).  
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Basommatophora;  
 OC Lymnaeidae; Planorbidae; Blomphalaria.  
 NCBI\_TaxID=6526;  
 RN [1]  
 RP SEQUENCE.  
 RC STRAIN=M-LINE; TISSUE=Hemolymph;  
 RX MEDLINE=97385165; PubMed=9238039;  
 RA Adema C.M., Herrel L.A., Miller R.D., Loker B.S.;  
 RT "A family of fibrinogen-related proteins that precipitates parasite-  
 derived molecules is produced by an invertebrate after infection.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:8691-8696(1997).  
 CC -1- FUNCTION: BINDS AND PRECIPITATES ANTIGENS OF THE PARASITE  
 CC ECHINOSTOMA PARASENSI.  
 CC -1- SUBCELLULAR LOCATION: SECRETED IN THE HEMOLYMPH.  
 CC -1- INDUCTION: By infection.  
 KM Lectin.  
 FT NON\_TER 1 16  
 FT NON\_TER 16  
 SQ SEQUENCE 16 AA; 1790 MW; 57489A8F2BEDA94 CRC64;  
 Query Match 19.2%; Score 19; DB 1; Length 16;  
 Best Local Similarity 37.5%; Pred. No. 4.3e+03;  
 Matches 3; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 DSIQWTD 9  
 DB 5 DLAQYVD 12  
 RESULT 14  
 TRYP\_PELCA STANDARD; PRT; 16 AA.  
 ID TRYP\_PELCA  
 AC P8107;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Trypsin precursor (EC 3.4.21.4) (Fragment).  
 OS Felis silvestris catus (Cat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Theria; Carnivora; Fissipedia; Felidae; Felis.  
 NCBI\_TaxID=9685;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=97235546; PubMed=9080665;  
 RA Steiner J.M., Medinger T.L., Williams D.A.;  
 RT "Purification and partial characterization of feline trypsin.";  
 RL Comp. Biochem. Physiol. 116B:87-93(1997).  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-Xaa, Lys-Xaa.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- SIMILARITY: BELONGS TO TRYPSIN FAMILY SL.  
 DR InterPro: IPR001254; Ser\_protease\_Try.  
 DR PROSITE: PS00240; TRYPSIN\_DOM; PARTIAL.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; PARTIAL.  
 DR PROSITE: PS00135; TRYPSIN\_SER; PARTIAL.  
 KM Hydrolyase; Serine protease; Digestion; Pancreas; Zymogen.  
 FT PROPEP 1 8  
 FT CHAIN 9  
 FT NON\_TER 9  
 FT NON\_TER 16  
 SQ SEQUENCE 16 AA; 1825 MW; A6D51BB58760A86 CRC64;  
 Query Match 19.2%; Score 19; DB 1; Length 16;  
 Best Local Similarity 60.0%; Pred. No. 4.3e+03;  
 Matches 3; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 14 ESDDK 18  
 DB 4 DDDDK 8

RESULT 15  
 LPM\_CORGL STANDARD; PRT; 17 AA.  
 ID LPM\_CORGL  
 AC P0656;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Trp operon leader peptide.  
 GN TRPL OR CGL3028.1.  
 OS Corynebacterium glutamicum (Brevibacterium flavum).  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.  
 NCBI\_TaxID=1718;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87117512; PubMed=3808947;  
 RA Matsui K., Sano K., Ohtsubo E.;  
 RT "Complete nucleotide and deduced amino acid sequences of the  
 RT Brevibacterium lactofermentum tryptophan operon.";  
 RL Nucleic Acids Res. 14:10113-10114(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88032866; PubMed=3607535;  
 RA Matsui K., Miwa K., Sano K.;  
 RT "Two single-base-pair substitutions causing desensitization to  
 RT tryptophan feedback inhibition of anthranilate synthase and enhanced  
 RT expression of tryptophan genes of Brevibacterium lactofermentum.";  
 RL J. Bacteriol. 169:5330-5332(1987).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87277409; PubMed=3609747;  
 RA Sano K., Matsui K.;  
 RT "Structure and function of the trp operon control regions of  
 RT Brevibacterium lactofermentum, a glutamic-acid-producing bacterium.";  
 RL Gene 53:191-200(1987).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=ATCC 13059 / AS019;  
 RX MEDLINE=91088299; PubMed=2263476;  
 RA Heery D.M., Duncan L.K.;  
 RT "Nucleotide sequence of the Corynebacterium glutamicum trp gene.";  
 RL Nucleic Acids Res. 18:7138-7138(1990).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;  
 RA Nakagawa S.;  
 RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN CONTROL OF THE BIOSYNTHESIS  
 CC OF TRYPTOPHAN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL: X04960; CAA28622.1; -;  
 DR EMBL: M17892; AAB59110.1; -;  
 DR EMBL: M16663; -; NOT\_ANNOTATED\_CDS.  
 DR EMBL: X55994; CAA39466.1; -;  
 DR EMBL: AP005283; -; NOT\_ANNOTATED\_CDS.  
 DR PIR: A29834; A29834.  
 KM Tryptophan biosynthesis; Leader peptide; Complete proteome.  
 SQ SEQUENCE 17 AA; 2112 MW; 74C7E7924DAAB56B CRC64;  
 Query Match 19.2%; Score 19; DB 1; Length 17;  
 Best Local Similarity 75.0%; Pred. No. 4.6e+03;  
 Matches 3; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 3 SLOW 6

Db 9 STOW 12

Search completed: January 21, 2004, 12:18:57  
Job time : 12 secs



```

DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Methylmalonyl CoA mutase (Fragment).
GN DR1189.
OS Thermus thermophilus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC Thermus.
OX NCB1_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KT8;
RA Spada S., Pembroke J.T., Wall J.G.;
RT "Cloning and characterisation of the czrB metal cation efflux protein
RT from T. thermophilus.";
RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ307316; CAC83723.1; -.
FT NON TER 16
SQ SEQUENCE 16 AA; 2176 MW; 423A2C2B2BA176C1 CRC64;

Query Match 25.3%; Score 25; DB 2; Length 16;
Best Local Similarity 37.5%; Pred. No. 3.5e+03;
Matches 3; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 6 WITDOYMK 13
DB 7 WLRRTYRK 14

RESULT 3
O53399 PRELIMINARY; PRT; 16 AA.
ID O53399
AC O53399;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE CRYIIB protein (Fragment).
GN CRYIIB.
OS Bacillus thuringiensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCB1_TaxID=1428;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94123945; PubMed=8293956;
RA Hoegman T.C., Ziniu Y., Shen J., Ellar D.J.;
RT "Identification of a cryptic gene associated with an insertion
RT sequence not previously identified in Bacillus thuringiensis.";
RL FEMS Microbiol. Lett. 114:23-29(1993).
DR EMBL; S68408; AAC60457.1; -.
FT NON TER 1
SQ SEQUENCE 16 AA; 1944 MW; 132368F185FD4F90 CRC64;

Query Match 25.3%; Score 25; DB 2; Length 16;
Best Local Similarity 57.1%; Pred. No. 3.5e+03;
Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 11 YNKESDD 17
DB 7 YNKESND 13

RESULT 4
P91578 PRELIMINARY; PRT; 14 AA.
ID P91578
AC P91578;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Nucleoside triphosphatase phosphohydrolase (Fragment).
GN NPH 1.
OS Choriactonura flumiferana entomopoxvirus (CEFPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae;
OC Entomopoxvirus B.

```

```

OX NCB1_TaxID=28322;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97239890; PubMed=9085546;
RA Li X., Barrett J.W., Yuen L., Arif B.M.;
RT "Cloning, sequencing and transcriptional analysis of the Choristoneura
RT flumiferana entomopoxvirus spheroidin gene.";
RL Virus Res. 47:143-154(1997).
DR EMBL; U19239; AAB39411.1; -.
FT NON TER 1
SQ SEQUENCE 14 AA; 1762 MW; 086D23ACF2B6650 CRC64;

Query Match 24.2%; Score 24; DB 12; Length 14;
Best Local Similarity 50.0%; Pred. No. 4.2e+03;
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 DSLQWITDOY 11
DB 4 DILYITNDY 13

RESULT 5
O81WX4 PRELIMINARY; PRT; 16 AA.
ID O81WX4
AC O81WX4;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Transcription factor AP-2 gamma (Fragment).
GN TFAP2G.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22378671; PubMed=12490322;
RA Li M., Wang Y., Yu Y., Nishizawa M., Nakajima T., Ito S., Kaman P.;
RT "The human transcription factor activation protein-2 gamma (AP-
RT 2gamma): gene structure, promoter, and expression in mammary carcinoma
RT cell lines.";
RL Gene 301:43-51(2002).
DR EMBL; AF544986; AAN37906.1; -.
FT NON TER 16
SQ SEQUENCE 16 AA; 2016 MW; F88CAEF56BFC601 CRC64;

Query Match 23.7%; Score 23.5; DB 4; Length 16;
Best Local Similarity 50.0%; Pred. No. 5.8e+03;
Matches 6; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

OY 6 W-ITDOYNKESD 16
DB 3 WKITDVKYSED 14

RESULT 6
Q9PYL3 PRELIMINARY; PRT; 14 AA.
ID Q9PYL3
AC Q9PYL3;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE Transcriptional transactivator Tax (Fragment).
GN TAX.
OS Human T-cell leukemia virus type II (HTLV-II).
OC Viruses; Retroid viruses; Retroviridae; Deltaretrovirus.
OX NCB1_TaxID=11909;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H2BC1;
RX MEDLINE=20014165; PubMed=10548129;
RA Peters A.A., Oger J.J., Coulchart M.B., Waters D.J., Cummings H.J.,

```

RA Dekaban G.A.;  
 RT "An apparent case of human T-cell lymphotropic virus type II (HTLV-  
 RT II)-associated neurological disease: a clinical, molecular, and  
 RT phylogenetic characterization."  
 DR J. Clin. Virol. 14:37-50(1999).  
 FT NON\_TER 1  
 SQ SEQUENCE 14 AA; 1567 MW; 7353D4CA8CC6041B CRC64;  
 Query Match 23.2%; Score 23; DB 15; Length 14;  
 Best Local Similarity 57.1%; Pred. No. 6e+03;  
 Matches 4; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 11 YNKSDD 17  
 Db 4 FNKEAD 10  
 RESULT 7  
 ID P79137 PRELIMINARY; PRT; 16 AA.  
 AC P79137;  
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)  
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)  
 DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)  
 DE Cyt1 region of membrane cofactor protein (MCP, CD46) (Fragment).  
 OS Cercopithecus aethiops (Green monkey) (Grivet).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Cercopithecus.  
 NC NCB1\_TaxID=9534;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=97149841; PubMed=8996635;  
 RA Murakami Y., Seya T., Kurita M., Nagasawa S.;  
 RT "Molecular cloning of a complementary DNA for a membrane cofactor  
 RT protein (MCP, CD46)/measles virus receptor on Vero cells and its  
 RT functional characterization."  
 RL Biol. Pharm. Bull. 19:1541-1545(1996).  
 DR EMBL: D78368; BAA11377.1; -.  
 FT NON\_TER 1  
 SQ SEQUENCE 16 AA; 1953 MW; 60C102BCD778F939 CRC64;  
 Query Match 23.2%; Score 23; DB 6; Length 16;  
 Best Local Similarity 33.3%; Pred. No. 6.9e+03;  
 Matches 3; Conservative 5; Mismatches 1; Indels 0; Gaps 0;  
 QY 6 WTDQNK 14  
 Db 2 YLTDEHRE 10  
 RESULT 8  
 ID Q8SPT4 PRELIMINARY; PRT; 16 AA.  
 AC Q8SPT4;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Chloride channel 2 (Fragment).  
 GN CLCN2.  
 OS Macaca mulatta (Rhesus macaque).  
 OC Eukaryota; Metazoa; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecinae; Macaca.  
 NC NCB1\_TaxID=9544;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Norgren R.B. Jr., Zink M.A., Jia Y., Ojeda S.R., Spindel E.R.;  
 RT "Construction of a targeted rhesus macaque microarray."  
 RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF489851; AAL96695.1; -.

FT NON\_TER 1  
 SQ SEQUENCE 16 AA; 1691 MW; 50D3BF4FC6AF49B CRC64;  
 Query Match 23.2%; Score 23; DB 6; Length 16;  
 Best Local Similarity 80.0%; Pred. No. 6.9e+03;  
 Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 14 ESDDK 18  
 Db 10 DSDDK 14  
 RESULT 9  
 ID Q9UR75 PRELIMINARY; PRT; 17 AA.  
 AC Q9UR75;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Ester hydrolase, CMEH (Fragment).  
 OS Yarrowia lipolytica (Candida lipolytica).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Dipodascaceae; Yarrowia.  
 NC NCB1\_TaxID=4952;  
 RN [1]  
 RP SEQUENCE  
 RX MEDLINE=95162576; PubMed=7858960;  
 RA Kallwas H.K., Yee C., Blythe T.A., McNabb T.J., Rogers E.E.,  
 RA Shames S.L.;  
 RT "Snzymes for the resolution of alpha-tertiary-substituted carboxylic  
 RT acid esters."  
 RL Bioorg. Med. Chem. 2:557-566(1994).  
 FT NON\_TER 17  
 SQ SEQUENCE 17 AA; 2104 MW; 5AA1D69D0C4B5C7 CRC64;  
 Query Match 23.2%; Score 23; DB 3; Length 17;  
 Best Local Similarity 42.9%; Pred. No. 7.4e+03;  
 Matches 3; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 5 QWTDQY 11  
 Db 4 RWISGDY 10  
 RESULT 10  
 ID Q6Z645 PRELIMINARY; PRT; 17 AA.  
 AC Q6Z645;  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Membrane cofactor protein CD46 (Fragment).  
 OS Saguinus oedipus (Cotton-top tamarin).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Saguinus.  
 NC NCB1\_TaxID=9490;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B95-8;  
 RX MEDLINE=96161817; PubMed=9494106;  
 RA Murakami Y., Seya T., Kurita M., Fukui A., Ueda S., Nagasawa S.;  
 RT "Molecular cloning of membrane cofactor protein (MCP, CD46) on B95a  
 RT cell, an Epstein-Barr virus-transformed marmoset B cell line: B95a-MCP  
 RT is susceptible to infection by the CAM, but not the Nagahata strain of  
 RT the measles virus."  
 RL Biochem. J. 330:1351-1359(1998).  
 DR EMBL: AB001991; BAA25631.1; -.  
 FT NON\_TER 1  
 SQ SEQUENCE 17 AA; 1983 MW; 1301A0ECD008F939 CRC64;  
 Query Match 23.2%; Score 23; DB 6; Length 17;  
 Best Local Similarity 33.3%; Pred. No. 7.4e+03;  
 Matches 3; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 6 WITDOYNKE 14  
: : : : :  
DB 3 YLDETHRE 11

## RESULT 11

Q30218 PRELIMINARY; PRT; 17 AA.  
ID Q30218  
AC Q30218;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE CD23 antigen (Fragment).  
GN FCER2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96124133; PubMed=8552454;  
RA Foster H.H., Masch R., Kretschmar T., Mischke D.,  
RA Ucanabka-Ziegler B., Ziegler A., Schmitt M., Wahn H.U.;  
RT "Genetic markers on chromosome 19p and prenatal diagnosis of HLA class  
RT II-deficient combined immunodeficiency.";  
RL Pediatr. Res. 38:812-816(1995).  
DR EMBL; S81114; AAB35925.2; -.  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 17 AA; 2076 MW; 53F5D4E75F1E5F47 CRC64;

Query Match 23.2%; Score 23; DB 7; Length 17;  
Best Local Similarity 30.8%; Pred. No. 7.4e+03;  
Matches 4; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 6 WITDOYNKESDDK 18  
: : : : :  
DB 1 WDTQSLKLEER 13

## RESULT 12

Q92Y75 PRELIMINARY; PRT; 18 AA.  
ID Q92Y75  
AC Q92Y75;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
DE Cytochrome oxidase II (Fragment).  
OS Cenocoeilus sp.  
OC Mitochondrion.  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preygota;  
OC Neoptera; Endopterygota; Hymenoptera; Apocrita; Ichneumonidea;  
OC Braconidae; Cenocoeilinae; Cenocoeilus.  
OX NCBI\_TaxID=81081;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99152621; PubMed=10028295;  
RA Dowton M., Austin A.D.;  
RT "Evolutionary dynamics of a mitochondrial rearrangement 'hot spot' in  
RT the Hymenoptera.";  
RL Mol. Biol. Evol. 16:298-309(1999).  
DR EMBL; AF082926; AAD1786.1; -.  
KM Mitochondrion.  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 18 AA; 2220 MW; ED42F77ADFB14386 CRC64;

Query Match 23.2%; Score 23; DB 8; Length 18;  
Best Local Similarity 50.0%; Pred. No. 7.9e+03;  
Matches 4; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 10 DPEFWITE 17

## RESULT 13

Q9QM82 PRELIMINARY; PRT; 18 AA.  
ID Q9QM82  
AC Q9QM82;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
DE PHOSPHOPHORIN=PEPTIDE Fragment 12-45 (Fragment).  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91291127; PubMed=2064607;  
RA Sabsay B., Stetler-Stevenson W.G., Lechner J.H., Vels A.;  
RT "Domain structure and sequence distribution in dentin phosphophoryn.";  
RL Biochem. J. 276:699-707(1991).  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 18 AA; 1993 MW; 8861C95493B5A5BA CRC64;

Query Match 23.2%; Score 23; DB 11; Length 18;  
Best Local Similarity 40.0%; Pred. No. 7.9e+03;  
Matches 4; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 8 TDQYNKESDD 17  
: : : : :  
DB 8 SDDSDSDSD 17

## RESULT 14

Q96QA7 PRELIMINARY; PRT; 10 AA.  
ID Q96QA7  
AC Q96QA7;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE BA151A2.1 (Cdc42 guanine exchange factor (GEF) 9 (Collipibistin, PEM-2,  
DE HPEM-2, KIAA0424) (Fragment).  
GN ARHGEP9.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Whitehead S.;  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL451106; CAC88408.1; -.  
FT NON\_TER 10  
FT NON\_TER 10  
SQ SEQUENCE 10 AA; 1122 MW; 39925CB878640043 CRC64;

Query Match 22.2%; Score 22; DB 4; Length 10;  
Best Local Similarity 75.0%; Pred. No. 6e+03;  
Matches 3; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 LOWI 7  
: : : : :  
DB 1 MOWI 4

## RESULT 15

Q9S8W8 PRELIMINARY; PRT; 15 AA.  
ID Q9S8W8  
AC Q9S8W8;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE SELF-Incompatibility S7 glycoprotein (Fragment).

OS Nicotiana glauca (winged tobacco) (Persian tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamiales; Solanales; Solanaceae; Nicotiana.  
 ON NCBI\_TaxId=4087;  
 RN [1]  
 RP SEQUENCE.  
 RX MEDLINE=92404717; PubMed=2535548;  
 RA Jahnke W., Batterham M.P., Clarke A.B., Moritz R.L., Simpson R.J.,  
 RT "Identification, isolation, and N-terminal sequencing of style  
 RT glycoproteins associated with self-incompatibility in Nicotiana  
 RT glauca";  
 RL Plant Cell 1:493-499 (1989).  
 DR InterPro; IPR015668; RNase\_T2.  
 DR Pfam; PF00445; ribonuclease\_T2; 1.  
 SQ SEQUENCE 15 AA; 1844 MW; D9F5CB05B215937 CRC64;

Query Match 22.2%; Score 22; DB 10; Length 15;  
 Best Local Similarity 80.0%; Pred. No. 9.1e+03;  
 Matches 4; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 LOWIT 8  
 Db 9 LOWIT 13

Search completed: January 21, 2004, 12:19:44  
 Job time : 35 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: January 21, 2004, 12:09:20 ; Search time 40 Seconds  
(without alignments)  
71.427 Million cell updates/sec

Title: US-09-941-314-7

Perfect score: 99  
Sequence: 1 KDSLQWITDQYNKESDDK 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 379592

Minimum DB seq length: 0  
Maximum DB seq length: 18

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

A. Geneseq.19Jun03:\*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	99	100.0	18	23	AAU79857
2	42	42.4	14	20	AAU78256
3	36	36.4	14	22	ABBS6674
4	34	34.3	14	19	AAW67541
5	34	34.3	14	22	AAW63778
6	34	34.3	14	23	ABBS7615
7	34	34.3	15	23	AAW51461
8	34	34.3	17	23	AAU88219
9	34	34.3	17	23	AAU90733

10	33	33.3	17	22	AAW63978	Amino acid sequenc
11	32	32.3	14	22	AAW67004	Human insulin-like
12	32	32.3	14	22	AAW67006	Human insulin-like
13	32	32.3	14	23	ABBS7703	IGFBP-1 binding pe
14	32	32.3	14	23	ABBS7705	IGFBP-1 binding pe
15	32	32.3	16	22	AAW62307	Domestic mite Bt5
16	32	32.3	16	23	AAW19318	Blomia tropicalis
17	32	32.3	17	23	AAU88216	Insulin/insulin-11
18	32	32.3	17	23	AAU90013	Insulin/insulin-11
19	31	31.3	13	23	AAW67542	Bos taurus isocitr
20	31	31.3	14	19	AAW67542	IGF-1/IGFBP inhibi
21	31	31.3	14	19	AAW67545	IGF-1/IGFBP inhibi
22	31	31.3	14	19	AAW67552	IGF-1/IGFBP inhibi
23	31	31.3	14	19	AAW67552	IGF-1/IGFBP inhibi
24	31	31.3	14	19	AAW67540	IGF-1/IGFBP inhibi
25	31	31.3	14	22	ABBS6760	Human SNP related
26	31	31.3	14	22	AAW63777	Synthetic peptide
27	31	31.3	14	22	AAW63779	Synthetic peptide
28	31	31.3	14	22	AAW63782	Synthetic peptide
29	31	31.3	14	22	AAW63783	Synthetic peptide
30	31	31.3	14	22	AAW63789	Synthetic peptide
31	31	31.3	14	22	AAW67005	Human insulin-like
32	31	31.3	14	22	AAW67007	Human insulin-like
33	31	31.3	14	22	AAW67008	Human insulin-like
34	31	31.3	14	22	AAW67037	Human insulin-like
35	31	31.3	14	22	AAW67038	Human insulin-like
36	31	31.3	14	23	ABBS7614	IGFBP-1 displacer
37	31	31.3	14	23	ABBS7619	IGFBP-1 displacer
38	31	31.3	14	23	ABBS7619	IGFBP-1 displacer
39	31	31.3	14	23	ABBS7620	IGFBP-1 displacer
40	31	31.3	14	23	ABBS7702	IGFBP-1 binding pe
41	31	31.3	14	23	ABBS7704	IGFBP-1 binding pe
42	31	31.3	14	23	ABBS7704	IGF/IGFBP interact
43	31	31.3	15	19	AAW67490	Synthetic peptide
44	31	31.3	15	22	AAW63724	Human insulin-like
45	31	31.3	15	22	AAW67041	Human insulin-like

## ALIGNMENTS

RESULT 1	
AAU79857	
AAU79857 standard; Peptide; 18 AA.	
AC	AAU79857;
DT	15-JUL-2002 (first entry)
XX	
XX	
DE	Human cystatin-8 (Zcys8) antigenic fragment #5.
XX	
KW	Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;
KW	spermatogenesis; seminal fluid viscosity; cryopreserved sperm;
KW	sperm motility; fertilisation; antigenic peptide.
OS	Homo sapiens.
XX	
PN	WO200220567-A2.
XX	
PD	14-MAR-2002.
XX	
XX	29-AUG-2001; 2001WO-US26868.
PP	
XX	01-SEP-2000; 2000US-230230P.
PR	
PA	(ZYMO) ZYMOGENETICS INC.
XX	
PI	Holloway JL, Gao Z, Bishop PD;
XX	
DR	WPI, 2002-383044/41.
XX	
PT	Novel isolated mammalian cystatin-8 polypeptide useful for promoting spermatogenesis, and inhibiting cancer procoagulant protein which leads

XX	PT	to inhibition of thrombotic events associated with cancer	-
XX	PS	Claim 2, Page 96, 100pp: English.	
XX	CC	The invention describes an isolated mammalian cystatin-8 (Zcys8)	
CC	CC	polypeptide (1). (1) is useful for: inhibiting cancer procoagulant	
CC	CC	protein in an individual and thus inhibiting the thrombotic events	
CC	CC	associated with cancer; promoting spermatogenesis, modulating seminal	
CC	CC	fluid viscosity, enhancing viability of cryopreserved sperm, sperm	
CC	CC	motility and fertilisation; and as antigenic peptides to generate	
CC	CC	antibodies. Zcys8 is useful as research reagent for characterising sites	
CC	CC	of interaction between Zcys8 and its receptor. Zcys8 is useful in	
CC	CC	enhancing fertilisation during assisted reproduction in humans and in	
CC	CC	animals. Anti-(1) antibodies are useful to screen biological samples like	
CC	CC	blood, urine, saliva, tissue biopsy and autopsy material in vitro for the	
CC	CC	presence of Zcys8. The antibodies are also useful to isolate large	
CC	CC	quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.	
CC	CC	The polynucleotide encoding (1) is useful to detect and to localise the	
CC	CC	expression of a Zcys8 gene in a biological sample and Zcys8	
CC	CC	oligonucleotide probes are useful for in vivo diagnosis. The	
CC	CC	polynucleotide encoding (1) is useful in determining whether a subject's	
CC	CC	chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene	
CC	CC	copy number changes, insertions, deletions, restriction site changes and	
CC	CC	rearrangements and genetic alterations that inactivate the Zcys8 gene.	
CC	CC	This sequence represents an antigenic peptide of human cystatin-8	
CC	CC	(Zcys8).	
XX	XX	Sequence 18 AA;	
SQ			
	Query Match	100.0%; Score 99; DB 23; Length 18;	
	Best Local Similarity	100.0%; Pred. No. 1.9e-08;	
	Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY		1 KDSLQWITDYQNKESDDK 18	
Db		1 KDSLQWITDYQNKESDDK 18	
RESULT 2			
AAW78256	ID	AAW78256 standard; Protein; 14 AA.	
XX	XX	AAW78256;	
AC	XX		
DT	XX	13-APR-1999 (first entry)	
DE	XX	Fragment of human secreted protein encoded by gene 28.	
XX	XX		
KM	KM	Human; secreted protein; fusion protein; gene therapy; protein therapy;	
KM	KM	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;	
KM	KM	developmental abnormality; foetal deficiency; blood; allergy; renal;	
KM	KM	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;	
KM	KM	inflammation; ischaemic shock; Alzheimer's disease; testostosis; AIDS;	
KM	KM	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;	
KM	KM	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;	
KM	KM	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.	
XX	OS	Homo sapiens.	
XX	XX	WO9856804-A1.	
PN	PN		
PD	XX	17-DEC-1998.	
XX	XX		
PF	XX	11-JUN-1998; 98WO-US12125.	
XX	XX		
PR	XX	02-OCT-1997; 97US-0061060.	
PR	PR	13-JUN-1997; 97US-0049547.	
PR	PR	13-JUN-1997; 97US-0049548.	
PR	PR	13-JUN-1997; 97US-0049549.	
PR	PR	13-JUN-1997; 97US-0049550.	
PR	PR	13-JUN-1997; 97US-0049566.	
PR	PR	13-JUN-1997; 97US-0049607.	
PR	PR	13-JUN-1997; 97US-0049608.	

XX		PR	13-JUN-1997;	97US-0049610.
XX		PR	13-JUN-1997;	97US-0049610.
XX		PR	13-JUN-1997;	97US-0049611.
XX		PR	13-JUN-1997;	97US-0050566.
XX		PR	13-JUN-1997;	97US-0050901.
XX		PR	13-JUN-1997;	97US-0052989.
XX		PR	08-JUL-1997;	97US-0051919.
XX		PR	18-AUG-1997;	97US-0055984.
XX		PR	12-SEP-1997;	97US-0058665.
XX		PR	12-SEP-1997;	97US-0058668.
XX		PR	12-SEP-1997;	97US-0058669.
XX		PR	12-SEP-1997;	97US-0058750.
XX		PR	12-SEP-1997;	97US-0058971.
XX		PR	12-SEP-1997;	97US-0058972.
XX		PR	12-SEP-1997;	97US-0058975.
XX		PR	02-OCT-1997;	97US-0060834.
XX		PR	02-OCT-1997;	97US-0060841.
XX		PR	02-OCT-1997;	97US-0060844.
XX		PR	02-OCT-1997;	97US-0060865.
XX		PR	02-OCT-1997;	97US-0061059.
XX		PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		PI	Brewer LA, Ebner R, Ferrie AM, Feng P, Greene JM, Laflaur DM;	
XX		PI	Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM, Shi Y, Young P,	
XX		PI	Yu GL;	
XX		DR	WPI; 1999-080881/07.	
XX		DR	N-PSTB; AAX04338.	
PT	New isolated human genes and the secreted polypeptides they encode -			
PT	useful for diagnosis and treatment of e.g. cancers, neurological			
PT	disorders, immune diseases, inflammation or blood disorders			
XX				
PS	Disclosure; Page 27; 380pp; English.			
CC	This sequence represents a fragment of a secreted human protein encoded			
CC	by the nucleic acid molecule detailed in the descriptor line. The gene			
CC	can be used to generate fusion proteins by linking to the gene to a			
CC	human immunoglobulin Fc portion (e.g. AAX04302) for increasing the			
CC	stability of the fused protein as compared to the human protein only.			
CC	The invention relates to 86 novel genes and their fragments (nucleic			
CC	acid sequences: AAX04311-X04410; amino acid sequences AAU78126-W78225)			
CC	which are useful for preventing, treating or ameliorating medical			
CC	conditions e.g. by protein or gene therapy. Also, pathological			
CC	conditions can be diagnosed by determining the amount of the new			
CC	polypeptides in a sample or by determining the presence of mutations in			
CC	the new polynucleotides. Specific uses are described for each of the 86			
CC	polynucleotides, based on which tissues they are most highly expressed in			
CC	(see AAX04311 for described uses).			
SQ	Sequence 14 AA;			
OY	Query Match 42.4%; Score 42; DB 20; Length 14;			
DG	Best Local Similarity 77.8%; Pred. No. 7.4;			
MATCHES	Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;			
OY	10 QYNKESDDK i8			
DG	:     :			
DG	1 EYNKESDDK 9			
RESULT 3				
ID	ABB56674 standard; Peptide; 14 AA.			
AC	ABB56674;			
DT	05-MAR-2002 (first entry)			
XX	Human SNP related amino acid sequence SEQ ID NO:1239.			
XX	Human; single nucleotide polymorphism; SNP; polymorphism; cytosstatic;			

KW immunosuppressive; antiinflammatory; neuroprotective; antimicrobial;  
 KW autoimmune disease; inflammation; cancer; nervous system disease;  
 KW infection; polymorphic protein.  
 XX Homo sapiens.  
 XX WO200138586-A2.  
 PN 31-MAY-2001.  
 XX 31-MAY-2001.  
 PD 22-NOV-2000; 2000MO-US32311.  
 XX 24-NOV-1999; 99US-0167383.  
 PR (CURA-) CURAGEN CORP.  
 XX Shinketsu RA, Leach M;  
 PI WPI; 2001-355949/37.  
 DR Isolated human nucleic acids comprising one or more single nucleotide  
 XX polymorphisms, useful for treating a subject suffering from a  
 PT pathology, e.g. autoimmune diseases, ascribed to the presence of a  
 PT sequence polymorphism -  
 XX Claim 1, Page 618; 674pp; English.  
 PS ABL00010 to ABL01104 represent human nucleic acid oligonucleotides  
 XX comprising one or more single nucleotide polymorphisms (SNPs). ABB56531  
 CC to ABB56503 represent human peptides encoded by some of the SNP  
 CC oligonucleotides. The sequences from the present invention can have  
 CC immunosuppressive, cytostatic, antiinflammatory, neuroprotective and  
 CC antimicrobial activities. Nucleic acids, polypeptides, oligonucleotides  
 CC and antibodies from the present invention can be used for treating a  
 CC subject suffering from, at risk for, or suspected of, suffering from a  
 CC pathology ascribed to the presence of a sequence polymorphism. The  
 CC pathology may be autoimmune diseases, inflammation, cancer, diseases of  
 CC the nervous system, and infection by pathogenic microorganisms. The SNPs  
 CC are also useful for determining which forms of a characterised  
 CC polymorphism are present in individuals. The antibodies may be used in  
 CC the detection, quantitation and/or cellular or tissue localisation of a  
 CC polymorphic protein (e.g., for use in measuring levels of the  
 CC polymorphic protein within appropriate physiological samples).  
 CC Sequence 14 AA;  
 SQ

Query Match 36.4%; Score 36; DB 22; Length 14;  
 Best Local Similarity 62.5%; Pred. No. 61;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 6 WITDOYRK 13  
 |||||:  
 5 WINDOYDR 12

RESULT 4  
 AAM67541  
 ID AAM67541 standard; peptide; 14 AA.  
 XX AAM67541;  
 AC 02-MAR-1999 (first entry)  
 DT IGF-1/IGFBP inhibitory peptide dpi-11.  
 XX IGF-1/IGFBP inhibitory peptide dpi-11.  
 DE Inhibition; interaction; insulin-like growth factor; IGF; receptor;  
 KW binding protein; serum; tissue; insulin; plasma; growth hormone;  
 KW glucose; secretion; blood; hyperglycaemia; obesity.  
 XX Synthetic.  
 OS WO9845427-A2.  
 XX

PD 15-OCT-1998.  
 XX 31-MAR-1998; 98MO-US06514.  
 PF 04-APR-1997; 97US-0825852.  
 PR (GETH ) GENENTECH INC.  
 XX Clark RG, Lowman HB, Robinson ICAF;  
 PI WPI; 1998-583196/49.  
 DR Inhibitors of interaction between insulin-like growth factor -  
 XX useful for, e.g. treating or preventing hyperglycaemia, obesity and  
 PT neurological disorders, and are optionally formulated with  
 PT thiazolidinone  
 XX Claim 21, Page 78; 133pp; English.  
 PS The invention relates to the isolation of compounds, especially  
 CC peptides, that: (a) inhibit interaction between an insulin-like growth  
 CC factor (IGF) with any one of its binding proteins (IGFBP), and (b) do not  
 CC bind to a human IGF receptor (hIGFR), e.g. see AAM67476-W67491 and  
 CC AAM67503-W67565. The peptides are produced synthetically or are isolated  
 CC from peptide libraries where the encoding sequence is generated so that  
 CC the resulting peptide produced retains its structure in solution. This  
 CC sequence represents a peptide which inhibits the interaction between  
 CC IGF-1 and IGFBP. The compounds are used to increase serum and tissue  
 CC levels of active IGF-1, preferably also to reduce insulin and plasma  
 CC growth hormone (GH) secretion, and blood glucose levels, without direct  
 CC stimulation of secretion or release of endogenous GH. Particularly the  
 CC compounds are used to treat or prevent hyperglycaemia, obesity-related  
 CC disorders, neurological, cardiac, anabolic, renal and immunological  
 CC diseases.  
 CC Sequence 14 AA;  
 SQ

Query Match 34.3%; Score 34; DB 19; Length 14;  
 Best Local Similarity 45.5%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 KDSLOWITDOY 11  
 |||||:  
 3 KGFLOWLCEKX 13

RESULT 5  
 AAG63778  
 ID AAG63778 standard; peptide; 14 AA.  
 XX AAG63778;  
 AC 29-OCT-2001 (first entry)  
 DT Synthetic peptide which binds to IGF-1 binding protein IGFBP-1.  
 XX Insulin-like growth factor; IGF; IGF-1; IGF binding protein; IGFBP-1;  
 KW hyperglycemic disorder; obesity-related disorder; neurological disorder;  
 KW cardiac disorder; anabolic disorder; renal disorder; neuroprotection;  
 KW immunological disorder; kidney regeneration; degenerative disorder;  
 KW hypoxia; wound healing; cardiac regeneration; cancer; angiogenesis;  
 KW metabolic stress; growth hormone deficiency; diabetes; short stature;  
 KW osteoporosis; obesity.  
 XX Synthetic.  
 OS US6251865-B1.  
 PN 26-JUN-2001.  
 PD 31-MAR-1998; 98US-0052888.  
 PF 04-APR-1997; 97US-0825852.  
 PR

```

XX (GERTH ) GENENTECH INC.
PA
XX
PI Clark RG, Lowman HB, Robinson ICAF;
XX
DR WPI; 2001-520042/57.
XX
PT Isolated peptides used to increase serum and tissue levels of
PT insulin-like growth factor in those with hyperglycemic,
PT obesity-related, neurological, cardiac, anabolic, renal or
PT immunological disorders
XX
PS Example 12; Column 77; 108bp; English.
XX
CC The present sequence represents a synthetic peptide, which binds to an
CC insulin-like growth factor (IGF)-1 binding protein. The specification
CC describes IGF agonists. IGF agonist peptides are used to increase
CC serum and tissue levels of IGF-1 in mammals with hyperglycemic,
CC obesity-related, neurological, cardiac, anabolic, renal or
CC immunological disorders. They may also be used to increase whole
CC body, bone and muscle growth rate in normal and hypopituitary animals,
CC to protect body weight and nitrogen loss during catabolic states,
CC kidney regeneration, to treat peripheral and central nervous system
CC (CNS) degenerative disorders and promote neuroprotection or repair
CC following CNS damage or injury, to treat hypoxia, to promote wound
CC healing, for cardiac regeneration, to reverse cancer cachexia, to
CC inhibit angiogenesis, to regenerate the gastrointestinal tract, to
CC stimulate mammary function, to counteract IGF-1-dependent actions of
CC growth hormone such as metabolic stress, age-related decline in growth
CC hormone activity and adult growth hormone deficiency, to treat maturity
CC onset diabetes and/or to treat specific IGF deficiency. They may also
CC be used to treat growth-hormone resistant short stature, growth hormone
CC insensitivity syndrome, osteoporosis and catabolic states, and reduce
CC obesity.
XX
SQ Sequence 14 AA;
XX
Query Match 34.3%; Score 34; DB 22; Length 14;
Best Local Similarity 45.5%; Pred. No. 1.2e+02;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 KDSLOWITDOY 11
Db 3 KGPIQWICEXY 13
XX
RESULT 6
ABBS7615
ID ABBS7615 standard; Peptide; 14 AA.
XX
AC ABBS7615;
XX
XX 18-MAR-2002 (first entry)
XX
DE IGFBP-1 displacer peptide #5.
XX
XX
KM Antineumatic; antiarthritic; osteopathic; cartilage disorder;
KM insulin-like growth factor; IGF; binding protein; IGFBP;
KM rheumatoid arthritis; osteoarthritis.
XX
OS Synthetic.
XX
XX
PN WO200187323-A2.
XX
PD 22-NOV-2001.
XX
XX
PF 16-MAY-2001; 2001WO-US15904.
XX
XX 16-MAY-2000; 2000US-204490P.
XX 15-NOV-2000; 2000US-248985P.
XX
PA (GERTH ) GENENTECH INC.
XX

```

```

PI Dubague Y, Filvaroff EH, Lowman HB;
XX
XX WPI; 2002-082942/11.
XX
DR
XX
PT Treating cartilage disorders including cartilage damage by injury or
PT degenerative cartilaginous disorders, by contacting cartilage with
PT insulin-like growth factor analog with altered affinity for IGF-binding
PT proteins
XX
XX Claim 13; Page 20; 136bp; English.
XX
XX
CC The present invention relates to a method for treating cartilage
CC disorders. The method comprises contacting cartilage with an active agent
CC such as insulin-like growth factor (IGF-1) analog with a binding affinity
CC preference for IGF binding protein-3 (IGFBP-3) over IGFBP-1, an IGF-1
CC analog with a binding affinity preference for IGFBP-1 over IGFBP-3, or a
CC IGFBP displacer peptide that prevents the interaction of IGF with an
CC IGFBP and does not bind to human IGF receptor. The method is useful for
CC treating cartilage disorders (CD), including degenerative CD, articular
CC CD such as rheumatoid arthritis and osteoarthritis. The present sequence
CC was used to illustrate the invention.
XX
SQ Sequence 14 AA;
XX
Query Match 34.3%; Score 34; DB 23; Length 14;
Best Local Similarity 45.5%; Pred. No. 1.2e+02;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 KDSLOWITDOY 11
Db 3 KGPIQWICEXY 13
XX
RESULT 7
AAM51461
ID AAM51461 standard; peptide; 15 AA.
XX
AC AAM51461;
XX
XX 25-JAN-2002 (first entry)
XX
DE Human CDV-1-42 peptide.
XX
XX Human; CDV-1-42; cytostatic; virucidal; immunomodulatory; HIV;
XX antiinflammatory; haemostatic; malignant tumour; gene therapy;
XX human immunodeficiency virus; infection; immunological disease.
XX
OS Homo sapiens.
XX
XX
PN WO200175061-A2.
XX
PD 11-OCT-2001.
XX
XX 26-MAR-2001; 2001WO-CN00534.
XX
XX 29-MAR-2000; 2000CN-0115255.
XX
PA (BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.
XX
XX Mao Y, Xie Y;
XX
XX WPI; 2002-025856/03.
XX
XX
XX CDV-1-42 and encoded polynucleotide, applicable in diagnosis and
XX treatment of malignant tumour, haemopathy, HIV infection, immunological
XX diseases and inflammation -
XX
XX Example 6; Page 13; 31bp; Chinese.
XX
XX The invention relates to human CDV-1-42 with cytostatic, virucidal,
XX immunomodulatory, antiinflammatory and haemostatic activity. The protein
XX and encoding polynucleotide are used in diagnosis and treatment of
XX malignant tumour, haemopathy, human immunodeficiency virus (HIV)
XX

```

CC infection, immunological diseases and various inflammations. The  
CC polynucleotide is useful in gene therapy. The present sequence is that of  
CC a peptide, useful to the invention.

XX Sequence 15 AA;

Query Match 34.3%; Score 34; DB 23; Length 15;  
Best Local Similarity 46.2%; Pred. No. 1.3e+02;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 2 DSIQWITDOYXKE 14  
| : : : | | | |  
Db 3 DQIKPTMDSLNKE 15

RESULT 8  
AAU88219 standard; Peptide; 17 AA.

XX AAU88219;  
XX 18-JUN-2002 (first entry)

DE Insulin/insulin-like growth factor receptor-binding peptide #183.

XX Cytostatic; antidiabetic; neuroprotective; cerebroprotective;  
XX ophthalmological; insulin; receptor; gene therapy; diabetes;  
XX insulin-like growth factor-1; IGF-1; tumour; prostate; breast;  
XX diabetic retinopathy; neurological diseases; stroke;  
XX diabetic neuropathy.

XX Synthetic.

XX WO200172771-A2.

XX 04-OCT-2001.

XX 29-MAR-2000; 2000MO-US08528.

XX 29-MAR-2000; 2000MO-US08528.

XX (DGIB-) DGI BIOTECHNOLOGIES LLC.

XX (NOVO) NOVO NORDISK AS.

XX Beasley J, Blume AJ, Schaeffer L, Pillutia R, Brandt J;  
PI Brissette R, Speitzler J, Cheng W, Ostergaard S, Mandeckl WS;  
PI Hansen PH, Ravera M, Hsiao K;

XX WPI; 2002-025774/03.

XX Modulating insulin activity in mammalian cells, for treating e.g.

PT diabetes and tumours, comprises using peptides that bind to insulin or

PT insulin-like growth factor receptors -

PS Disclosure; Page 44; 390pp; English.

XX The invention relates to a method of modulating insulin activity in  
CC mammalian cells by administering a peptide that binds the insulin  
CC receptor (IR). A composition containing a peptide, optionally expressed  
CC from gene therapy vectors, that binds to site 1 of IR and an insulin  
CC agonist are useful for treating diabetes. Also, peptides that are  
CC antagonists of the insulin-like growth factor-1 (IGF-1) receptor are  
CC useful for treating insulin-like growth factor (IGF)-sensitive tumours  
CC (e.g. of prostate and breast) and diabetic retinopathy, while IGF-1  
CC receptor agonists are useful for treating neurological diseases,  
CC including stroke and diabetic neuropathy. The peptides are also useful in  
CC screening for compounds that bind to IR or IGF-1 receptor, potential  
CC therapeutics and research reagents. AAU88034-AAU90957 represent IR  
CC and/or IGF-1 receptor-binding peptides and related amino acid sequences  
XX of the invention.

XX Sequence 17 AA;

Query Match 34.3%; Score 34; DB 23; Length 17;  
Best Local Similarity 62.5%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 6 WITDOYXK 13  
| : : : | | | |  
Db 10 WPAQDYXK 17

RESULT 9  
AAU90733 standard; Peptide; 17 AA.

XX AAU90733;

XX 18-JUN-2002 (first entry)

DE Insulin/insulin-like growth factor receptor-binding peptide #2689.

XX Cytostatic; antidiabetic; neuroprotective; cerebroprotective;  
XX ophthalmological; insulin; receptor; gene therapy; diabetes;  
XX insulin-like growth factor-1; IGF-1; tumour; prostate; breast;  
XX diabetic retinopathy; neurological diseases; stroke;  
XX diabetic neuropathy.

XX Synthetic.

XX WO200172771-A2.

XX 04-OCT-2001.

XX 29-MAR-2000; 2000MO-US08528.

XX 29-MAR-2000; 2000MO-US08528.

XX (DGIB-) DGI BIOTECHNOLOGIES LLC.

XX (NOVO) NOVO NORDISK AS.

XX Beasley J, Blume AJ, Schaeffer L, Pillutia R, Brandt J;  
PI Brissette R, Speitzler J, Cheng W, Ostergaard S, Mandeckl WS;  
PI Hansen PH, Ravera M, Hsiao K;

XX WPI; 2002-025774/03.

XX Modulating insulin activity in mammalian cells, for treating e.g.

PT diabetes and tumours, comprises using peptides that bind to insulin or

PT insulin-like growth factor receptors -

PS Disclosure; Figure 11B; 390pp; English.

XX The invention relates to a method of modulating insulin activity in  
CC mammalian cells by administering a peptide that binds the insulin  
CC receptor (IR). A composition containing a peptide, optionally expressed  
CC from gene therapy vectors, that binds to site 1 of IR and an insulin  
CC agonist are useful for treating diabetes. Also, peptides that are  
CC antagonists of the insulin-like growth factor-1 (IGF-1) receptor are  
CC useful for treating insulin-like growth factor (IGF)-sensitive tumours  
CC (e.g. of prostate and breast) and diabetic retinopathy, while IGF-1  
CC receptor agonists are useful for treating neurological diseases,  
CC including stroke and diabetic neuropathy. The peptides are also useful in  
CC screening for compounds that bind to IR or IGF-1 receptor, potential  
CC therapeutics and research reagents. AAU88034-AAU90957 represent IR  
CC and/or IGF-1 receptor-binding peptides and related amino acid sequences  
XX of the invention.

XX Sequence 17 AA;

Query Match 34.3%; Score 34; DB 23; Length 17;  
Best Local Similarity 62.5%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 6 WITDOYXK 13  
| : : : | | | |

Db	10	WFADQYKK 17
RESULT 10		
ID	AAG63978	standard; peptide; 17 AA.
AC	AAG63978;	
XX	13-NOV-2001	(first entry)
DT	13-NOV-2001	(first entry)
XX	Amino acid sequence of recombinant NS1 proteins.	
DE	Nonstructural gene; NS gene; influenza A virus; NS1 gene; vaccine;	
XX	viral infection; influenza infection; HIV-1 infection; gp41.	
KW	Synthetic.	
XX	WO200164860-A2.	
XX	07-SEP-2001.	
XX	02-MAR-2001; 2001WO-EP02392.	
PF	02-MAR-2000; 2000EP-0104338.	
XX	(POLY-) POLYMER SCI IMMUNOBIOLOGISCHE FORSCHUNG.	
PA	Ferko B, Egorov A, Voglauer R;	
XX	WPI, 2001-514840/56.	
DR	Recombinant NS gene of an influenza A virus comprising a functional RNA	
XX	binding domain and a gene sequence modification after nucleotide	
FT	position 400 of the NS1 gene segment, useful for producing a live	
PT	attenuated influenza virus vaccine -	
XX		
XX	Example 1; Fig 2; 40pp; English.	
XX	The specification describes a recombinant nonstructural (NS) gene of	
CC	an influenza A virus. The gene comprises a functional RNA binding	
CC	domain and a gene sequence modification after nucleotide position 400	
CC	of the NS1 gene segment, counted on the basis of influenza A/PR/8/34	
CC	virus, where the modification bars transcription of the remaining	
CC	portion of the NS1 gene segment. The recombinant NS gene is used to	
CC	produce a vaccine, which is useful for prophylactic or therapeutic	
CC	application against a viral infection, preferably against influenza	
CC	or HIV-1 infection. Influenza virus transfectants that contain the	
CC	modified NS gene may have an interferon (IFN) inducing phenotype, but	
CC	may or may not be sensitive towards to IFN. The present sequence	
CC	represents a fragment of a recombinant NS protein of the invention,	
CC	comprising gp41 and IL-1beta peptides.	
XX		
SO	Sequence	17 AA;
Query Match	33.3%;	Score 33; DB 22; Length 17;
Best Local Similarity	42.9%;	Pred. No. 2.2e+02;
Matches	6; Conservative	4; Mismatches 2; Indels 0; Gaps 0;
OY	5 QWTDQYKESDDK 18	
DB	4 KMASVQGEESNDK 17	
RESULT 11		
ID	AAB67004	standard; Peptide; 14 AA.
AC	AAB67004;	
XX	09-APR-2001	(first entry)
DT	09-APR-2001	(first entry)
XX	Human insulin-like growth factor agonist peptide SEQ ID NO: 4.	
DE		

```

XX Human; insulin-like growth factor; IGF; agonist; hyperglycaemic disorder;
KW obesity; neurological disorder; cardiac disorder; renal disorder;
KW immunological disorder; anabolic disorder.
XX
OS Homo sapiens.
XX
PN WO200078801-A2.
XX
PD 28-DEC-2000.
XX
PF 19-JUN-2000; 2000WO-US17023.
XX
PR 22-JUN-1999; 99US-0337227.
XX
PA (GETH ) GENENTECH INC.
XX
PI Chen YM, Cochran AG, Lowman HB, Skelton NJ;
XX
DR WPI; 2001-112312/12.
XX
PT New peptide for increasing serum and tissue levels of biological active
XX
PS Claim 18; Page 8; 75pp; English.
XX
CC The present invention provides the sequences of a number of peptides
CC which act as human insulin growth factor (IGF) agonists. These can be
CC used in the treatment of hyperglycaemic, obesity-related, neurological,
CC cardiac, renal, immunological and anabolic disorders.
XX
SQ Sequence 14 AA,
XX
Query Match 32.3%; Score 32; DB 22; Length 14;
Best Local Similarity 44.4%; Pred. No. 2.5e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 3 SLOWITDOX 11
:|:|:|:|
Db 5 ALQWLCEKY 13

```

PT insulin growth factor -  
 XX  
 PS Claim 18; Page 8; 75pp; English.  
 XX  
 CC The present invention provides the sequences of a number of peptides  
 CC which act as human insulin growth factor (IGF) agonists. These can be  
 CC used in the treatment of hyperglycaemic, obesity-related, neurological,  
 CC cardiac, renal, immunological and anabolic disorders.  
 XX  
 SQ Sequence 14 AA;  
 XX  
 Query Match 32.3%; Score 32; DB 23; Length 14;  
 Best Local Similarity 44.4%; Pred. No. 2.5e+02;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 QY 3 SLOWITDOY 11  
 :|||: :||  
 Db 5 NLOWLCEKY 13  
 XX  
 RESULT 13  
 ABB57703  
 ID ABB57703 standard; Peptide; 14 AA.  
 AC ABB57703;  
 XX  
 DT 18-MAR-2002 (first entry)  
 XX  
 DE IGFBP-1 binding peptide #2.  
 XX  
 KW Antirheumatic; antiarthritic; osteopathic; cartilage disorder;  
 KW insulin-like growth factor; IGF; binding protein; IGFBP;  
 KW rheumatoid arthritis; osteoarthritis.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200187323-A2.  
 XX  
 PD 22-NOV-2001.  
 XX  
 PF 16-MAY-2001; 2001WO-US15904.  
 XX  
 PR 16-MAY-2000; 2000US-204490P.  
 PR 15-NOV-2000; 2000US-248985P.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Dubaquié Y, Flivarooff EH, Lowman HB;  
 XX  
 DR WPI; 2002-082942/11.  
 XX  
 PT Treating cartilage disorders including cartilage damage by injury or  
 PT degenerative cartilaginous disorders, by contacting cartilage with  
 PT insulin-like growth factor analog with altered affinity for IGF-binding  
 PT proteins -  
 XX  
 PS Example 12; Page 87; 136pp; English.  
 XX  
 CC The present invention relates to a method for treating cartilage  
 CC disorders. The method comprises contacting cartilage with an active agent  
 CC such as insulin-like growth factor (IGF-1) analog with a binding affinity  
 CC preference for IGF binding protein-3 (IGFBP-3) over IGFBP-1, an IGF-1  
 CC analog with a binding affinity preference for IGFBP-1 over IGFBP-3, or a  
 CC IGFBP displacer peptide that prevents the interaction of IGF with an  
 CC IGFBP and does not bind to human IGF receptor. The method is useful for  
 CC treating cartilage disorders (CD), including degenerative CD, articular  
 CC CD such as rheumatoid arthritis and osteoarthritis. The present sequence  
 CC was used to illustrate the invention.  
 XX  
 SQ Sequence 14 AA;  
 XX  
 Query Match 32.3%; Score 32; DB 23; Length 14;  
 Best Local Similarity 44.4%; Pred. No. 2.5e+02;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 QY 3 SLOWITDOY 11  
 :|||: :||  
 Db 5 ALOWLCEKY 13  
 XX  
 RESULT 14  
 ABB57705  
 ID ABB57705 standard; Peptide; 14 AA.  
 AC ABB57705;  
 XX  
 DT 18-MAR-2002 (first entry)  
 XX  
 DE IGFBP-1 binding peptide #4.  
 XX  
 KW Antirheumatic; antiarthritic; osteopathic; cartilage disorder;  
 KW insulin-like growth factor; IGF; binding protein; IGFBP;  
 KW rheumatoid arthritis; osteoarthritis.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200187323-A2.  
 XX  
 PD 22-NOV-2001.  
 XX  
 PF 16-MAY-2001; 2001WO-US15904.  
 XX  
 PR 16-MAY-2000; 2000US-204490P.  
 PR 15-NOV-2000; 2000US-248985P.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Dubaquié Y, Flivarooff EH, Lowman HB;  
 XX  
 DR WPI; 2002-082942/11.  
 XX  
 PT Treating cartilage disorders including cartilage damage by injury or  
 PT degenerative cartilaginous disorders, by contacting cartilage with  
 PT insulin-like growth factor analog with altered affinity for IGF-binding  
 PT proteins -  
 XX  
 PS Example 12; Page 87; 136pp; English.  
 XX  
 CC The present invention relates to a method for treating cartilage  
 CC disorders. The method comprises contacting cartilage with an active agent  
 CC such as insulin-like growth factor (IGF-1) analog with a binding affinity  
 CC preference for IGF binding protein-3 (IGFBP-3) over IGFBP-1, an IGF-1  
 CC analog with a binding affinity preference for IGFBP-1 over IGFBP-3, or a  
 CC IGFBP displacer peptide that prevents the interaction of IGF with an  
 CC IGFBP and does not bind to human IGF receptor. The method is useful for  
 CC treating cartilage disorders (CD), including degenerative CD, articular  
 CC CD such as rheumatoid arthritis and osteoarthritis. The present sequence  
 CC was used to illustrate the invention.  
 XX  
 SQ Sequence 14 AA;  
 XX  
 Query Match 32.3%; Score 32; DB 23; Length 14;  
 Best Local Similarity 44.4%; Pred. No. 2.5e+02;  
 Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
 QY 3 SLOWITDOY 11  
 :|||: :||  
 Db 5 NLOWLCEKY 13  
 XX  
 RESULT 15  
 AAB02307  
 ID AAB02307 standard; peptide; 16 AA.  
 AC AAB02307;  
 XX

DT 31-JUL-2001 (first entry)  
 XX Domestic mite Bt5 allergen peptide variant, P83.  
 DE  
 XX  
 KM Mite; immunogenic protein; Bt allergen; therapy; atopic dermatitis;  
 KW immediate hypersensitivity; systemic anaphylaxis; allergic rhinitis;  
 KM asthma; anti-allergic; anti-inflammatory; immunosuppressive; variant.  
 XX  
 OS Blomia tropicalis.  
 XX  
 PN W0200130817-A1.  
 XX  
 PD 03-MAY-2001.  
 XX  
 PF 10-OCT-2000; 2000WO-AU01227.  
 XX  
 PR 26-OCT-1999; 99SG-0005313.  
 PR 18-JUL-2000; 2000AU-0008842.  
 PR 18-JUL-2000; 2000AU-0008844.  
 PR 18-JUL-2000; 2000AU-0008845.  
 XX  
 PA (UYSI-) UNIV SINGAPORE NAT.  
 XX  
 PI Chua KY, Cheong N, Lee BW;  
 XX  
 DR WPI; 2001-308609/32.  
 XX  
 PT Novel immunogenic protein derived from house mite, Blomia tropicalis  
 PT useful for treating and diagnosing conditions involving induction of  
 PT immuneresponse to mite, such as allergic asthma, atopic dermatitis,  
 PT rhinitis  
 XX  
 PS Example 43; Fig 20; 230bp; English.  
 XX  
 CC The present invention relates to immunogenic proteins, referred as Bt  
 CC allergen, is derived from domestic mite, Blomia tropicalis. The specific  
 CC Bt allergens of the invention includes Bt1, Bt10, Bt5 and BtA2. The  
 CC immunogenic protein is useful for preventing, reducing or ameliorating  
 CC Blomia tropicalis hypersensitivity condition such as atopic dermatitis,  
 CC immediate hypersensitivity, systemic anaphylaxis, allergic rhinitis or  
 CC asthma and for modulating an immune response directed to Bt allergen in  
 CC a subject. The Bt allergens are also useful for detecting antibody  
 CC directed to all or a part of Bt allergen in a biological sample from a  
 CC subject. Antibodies to Bt allergens are also used as therapeutic or  
 CC diagnostic agents, to screen Bt immunoassays and as antagonists to  
 CC inhibit Bt activity under circumstances where temporary hypersensitivity  
 CC inhibition is required. The present sequence is Bt5 allergen  
 CC peptide variant.  
 CC  
 SQ Sequence 16 AA;  
 Query March 32.3%; Score 32; DB 22; Length 16;  
 Best Local Similarity 57.1%; Pred. No. 2.9e+02;  
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 6 WIRDOYN 12  
 | : |||  
 Db 8 WLRQYN 14

Search completed: January 21, 2004, 12:18:33  
 Job time : 41 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 12:19:51 ; Search time 31 Seconds  
(without alignments)  
118.731 Million cell updates/sec

Title: US-09-941-314-7

Perfect score: 99

Sequence: 1 KDSLOWITDQYNKESDDK 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 762491 seqs, 204481190 residues

Total number of hits satisfying chosen parameters: 157203

Minimum DB seq length: 0

Maximum DB seq length: 18

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10E\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99	100.0	18	10	US-09-941-314-7
2	34	34.3	14	11	US-09-858-935B-30
3	34	34.3	17	12	US-09-962-756-1200
4	34	34.3	17	12	US-09-962-756-1756
5	34	34.3	17	12	US-10-253-471-1200
6	34	34.3	17	12	US-10-253-471-1756
7	33	33.3	14	15	US-10-283-838-12
8	32	32.3	14	11	US-09-858-935B-132
9	32	32.3	14	11	US-09-858-935B-134
10	32	32.3	17	12	US-09-962-756-1752
11	32	32.3	17	12	US-10-253-471-1752
12	31	31.3	12	11	US-09-884-696-8
13	31	31.3	13	10	US-09-897-107-81
14	31	31.3	14	11	US-09-858-935B-29
15	31	31.3	14	11	US-09-858-935B-31

16	31	31.3	14	11	US-09-858-935B-34	Sequence 34, Appl
17	31	31.3	14	11	US-09-858-935B-35	Sequence 35, Appl
18	31	31.3	14	11	US-09-858-935B-44	Sequence 44, Appl
19	31	31.3	14	11	US-09-858-935B-131	Sequence 131, Appl
20	31	31.3	14	11	US-09-858-935B-133	Sequence 133, Appl
21	31	31.3	15	11	US-09-858-935B-26	Sequence 26, Appl
22	31	31.3	16	15	US-10-186-867-57	Sequence 57, Appl
23	31	31.3	17	11	US-09-858-935B-41	Sequence 41, Appl
24	31	31.3	18	11	US-09-858-935B-39	Sequence 39, Appl
25	30	30.3	14	9	US-09-765-527-133	Sequence 133, App
26	30	30.3	14	9	US-09-881-490-104	Sequence 104, App
27	30	30.3	16	15	US-10-225-322-15	Sequence 15, Appl
28	30	30.3	17	15	US-10-225-322-4	Sequence 4, Appl
29	29	29.3	9	10	US-09-779-308-411	Sequence 411, App
30	29	29.3	10	10	US-09-779-308-319	Sequence 319, App
31	29	29.3	10	11	US-09-858-935B-144	Sequence 144, App
32	29	29.3	10	11	US-09-858-935B-146	Sequence 146, App
33	29	29.3	11	11	US-09-858-935B-145	Sequence 145, App
34	29	29.3	12	11	US-09-884-696-10	Sequence 10, Appl
35	29	29.3	12	11	US-09-858-935B-53	Sequence 53, Appl
36	29	29.3	14	11	US-09-858-935B-52	Sequence 52, Appl
37	29	29.3	14	11	US-09-858-935B-54	Sequence 54, Appl
38	29	29.3	16	12	US-09-825-517A-135	Sequence 135, Appl
39	29	29.3	16	15	US-10-293-822-7	Sequence 7, Appl
40	29	29.3	17	15	US-10-225-567A-713	Sequence 713, App
41	28	28.8	18	12	US-10-145-206-168	Sequence 168, App
42	28	28.3	7	12	US-10-083-894-9	Sequence 9, Appl
43	28	28.3	14	11	US-09-858-935B-138	Sequence 138, App
44	28	28.3	17	15	US-10-094-401-201	Sequence 201, App
45	28	28.3	17	15	US-10-272-411-32	Sequence 32, Appl

#### ALIGNMENTS

RESULT 1  
US-09-941-314-7  
; Sequence 7, Application US/09941314  
; Patent No. US20020142396A1  
; GENERAL INFORMATION:  
; APPLICANT: ZymoGenetics, Inc.  
; TITLE OF INVENTION: Mammalian Cystatin-B and Its Use to  
; TITLE OF INVENTION: Inhibit Cancer Procoagulant Protein  
; FILE REFERENCE: 00-81PC  
; CURRENT APPLICATION NUMBER: US/09/941,314  
; CURRENT FILING DATE: 2001-08-29  
; PRIOR APPLICATION NUMBER: 60/230,230  
; PRIOR FILING DATE: 2001-09-01  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-941-314-7

Query Match 100.0%; Score 99; DB 10; Length 18;  
Best Local Similarity 100.0%; Pred. No. 6.7e-08;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KDSLOWITDQYNKESDDK 18  
Db 1 KDSLOWITDQYNKESDDK 18

RESULT 2  
US-09-858-935B-30  
; Sequence 30, Application US/09858935B  
; Publication No. US20030069177A1  
; GENERAL INFORMATION:  
; APPLICANT: Dubaque, Yves  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Lowman, Henry B.

;; TITLE OF INVENTION: METHOD FOR TREATING CARTILAGE DISORDERS  
;; FILE REFERENCE: P1794R1  
;; CURRENT APPLICATION NUMBER: US/09/858,935B  
;; CURRENT FILING DATE: 2002-07-02  
;; PRIOR APPLICATION NUMBER: US 60/248,985  
;; PRIOR FILING DATE: 2000-11-15  
;; PRIOR APPLICATION NUMBER: US 60/204,490  
;; PRIOR FILING DATE: 2000-05-16  
;; NUMBER OF SEQ ID NOS: 153  
;; SEQ ID NO 30  
;; LENGTH: 14  
;; TYPE: PRT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Sequence is synthesized  
US-09-858-935B-30

Query Match 34.3%; Score 34; DB 11; Length 14;  
Best Local Similarity 45.5%; Pred. No. 1.5e+02;  
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 KDSQWITDOY 11  
| | | | |  
Db 3 KGPLQWLCEKY 13

RESULT 3  
US-09-962-756-1200  
;; Sequence 1200, Application US/09962756  
;; Publication No. US20030195147A1  
;; GENERAL INFORMATION:  
;; APPLICANT: PILUTLA, RENUKA  
;; APPLICANT: BRISSETTE, RENEE  
;; APPLICANT: BLUME, ARTHUR J.  
;; APPLICANT: SCHAEFER, LANGE  
;; APPLICANT: BRANDT, JAKOB  
;; APPLICANT: GOLDSTEIN, NEIL I.  
;; APPLICANT: SPETZLER, JANE  
;; APPLICANT: OSTERGAARD, SOREN  
;; APPLICANT: HANSEN, PER HERTZ  
;; TITLE OF INVENTION: INSULIN AND IGF-1 RECEPTOR AGONISTS AND ANTAGONISTS  
;; FILE REFERENCE: 1878-4051US1  
;; CURRENT APPLICATION NUMBER: US/09/962,756  
;; CURRENT FILING DATE: 2001-09-24  
;; PRIOR APPLICATION NUMBER: 09/538,038  
;; PRIOR FILING DATE: 2000-03-29  
;; PRIOR APPLICATION NUMBER: 09/146,127  
;; PRIOR FILING DATE: 1998-09-02  
;; NUMBER OF SEQ ID NOS: 2227  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 1200  
;; LENGTH: 17  
;; TYPE: PRT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
;; OTHER INFORMATION: peptide  
US-09-962-756-1200

Query Match 34.3%; Score 34; DB 12; Length 17;  
Best Local Similarity 62.5%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 WITDOYRK 13  
| | | | |  
Db 10 WPADQYRK 17

RESULT 4  
US-09-962-756-1756  
;; Sequence 1756, Application US/09962756  
;; Publication No. US20030195147A1  
;; GENERAL INFORMATION:

;; APPLICANT: PILUTLA, RENUKA  
;; APPLICANT: BRISSETTE, RENEE  
;; APPLICANT: BLUME, ARTHUR J.  
;; APPLICANT: SCHAEFER, LANGE  
;; APPLICANT: BRANDT, JAKOB  
;; APPLICANT: GOLDSTEIN, NEIL I.  
;; APPLICANT: SPETZLER, JANE  
;; APPLICANT: OSTERGAARD, SOREN  
;; APPLICANT: HANSEN, PER HERTZ  
;; TITLE OF INVENTION: INSULIN AND IGF-1 RECEPTOR AGONISTS AND ANTAGONISTS  
;; FILE REFERENCE: 1878-4051US1  
;; CURRENT APPLICATION NUMBER: US/09/962,756  
;; CURRENT FILING DATE: 2001-09-24  
;; PRIOR APPLICATION NUMBER: 09/538,038  
;; PRIOR FILING DATE: 2000-03-29  
;; PRIOR APPLICATION NUMBER: 09/146,127  
;; PRIOR FILING DATE: 1998-09-02  
;; NUMBER OF SEQ ID NOS: 2227  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 1756  
;; LENGTH: 17  
;; TYPE: PRT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
;; OTHER INFORMATION: peptide  
US-09-962-756-1756

Query Match 34.3%; Score 34; DB 12; Length 17;  
Best Local Similarity 62.5%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 WITDOYRK 13  
| | | | |  
Db 10 WPADQYRK 17

RESULT 5  
US-10-253-471-1200  
;; Sequence 1200, Application US/10253471  
;; Publication No. US20030236190A1  
;; GENERAL INFORMATION:  
;; APPLICANT: PILUTLA, RENUKA et al.  
;; TITLE OF INVENTION: INSULIN AND IGF-1 RECEPTOR AGONISTS AND ANTAGONISTS  
;; FILE REFERENCE: 1878-4057  
;; CURRENT APPLICATION NUMBER: US/10/253,471  
;; CURRENT FILING DATE: 2002-09-24  
;; PRIOR APPLICATION NUMBER: 09/962,756  
;; PRIOR FILING DATE: 2001-09-24  
;; PRIOR APPLICATION NUMBER: 09/538,038  
;; PRIOR FILING DATE: 2000-03-29  
;; PRIOR APPLICATION NUMBER: 09/146,127  
;; PRIOR FILING DATE: 1998-09-02  
;; NUMBER OF SEQ ID NOS: 2227  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 1200  
;; LENGTH: 17  
;; TYPE: PRT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
;; OTHER INFORMATION: peptide  
US-10-253-471-1200

Query Match 34.3%; Score 34; DB 12; Length 17;  
Best Local Similarity 62.5%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 WITDOYRK 13  
| | | | |  
Db 10 WPADQYRK 17

RESULT 6  
US-10-253-471-1756  
Sequence 1756, Application US/10253471  
Publication No. US20030236190A1  
GENERAL INFORMATION:  
APPLICANT: PILITULA, RENUKA et al.  
TITLE OF INVENTION: INSULIN AND IGF-1 RECEPTOR AGONISTS AND ANTAGONISTS  
FILE REFERENCE: 1878-4057  
CURRENT APPLICATION NUMBER: US/10/253,471  
PRIOR FILING DATE: 2002-09-24  
PRIOR APPLICATION NUMBER: 09/962,756  
PRIOR FILING DATE: 2001-09-24  
PRIOR APPLICATION NUMBER: 09/538,038  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: 09/146,127  
PRIOR FILING DATE: 1998-09-02  
NUMBER OF SEQ ID NOS: 2227  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 1756  
LENGTH: 17  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-10-253-471-1756

Query Match 34.3%; Score 34; DB 12; Length 17;  
Best Local Similarity 62.5%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 WITDQYK 13  
DB 10 WPADQYK 17

RESULT 7  
US-10-283-838-12  
Sequence 12, Application US/10283838  
Publication No. US20030092894A1  
GENERAL INFORMATION:  
APPLICANT: Per Antonsson, Per Bjork, Mikael Dohlsten,  
Johan Hansson, Terje Kalland, Lars  
Abrahmsen and Goran Forsberg  
TITLE OF INVENTION: MODIFIED/CHIMERIC SUPRANATIGENS  
AND THEIR USE  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pravel, Hewitt, Kimball & Krieger  
STREET: 1177 West Loop South, 10th Floor  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77027-9095  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/283,838  
FILING DATE: 30-Oct-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/695,692  
FILING DATE: August 12, 1996  
APPLICATION NUMBER: 9601245-5  
FILING DATE: March 29, 1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Krieger, Paul B.  
REGISTRATION NUMBER: 25,886  
REFERENCE/DOCKET NUMBER: 41986/1  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 713-850-0909  
TELEFAX: 713-850-0165  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
US-10-283-838-12

Query Match 33.3%; Score 33; DB 15; Length 14;  
Best Local Similarity 85.7%; Pred. No. 2.1e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 12 NKESDDK 18  
DB 4 NKESDDQ 10

RESULT 8  
US-09-858-935B-132  
Sequence 132, Application US/09858935B  
Publication No. US20030069177A1  
GENERAL INFORMATION:  
APPLICANT: Dubague, Yves  
APPLICANT: Lowman, Henry B.  
TITLE OF INVENTION: METHOD FOR TREATING CARTILAGE DISORDERS  
FILE REFERENCE: P1794R1  
CURRENT APPLICATION NUMBER: US/09/858,935B  
CURRENT FILING DATE: 2002-07-02  
PRIOR APPLICATION NUMBER: US 60/248,985  
PRIOR FILING DATE: 2000-11-15  
PRIOR APPLICATION NUMBER: US 60/204,490  
PRIOR FILING DATE: 2000-05-16  
NUMBER OF SEQ ID NOS: 153  
SEQ ID NO 132  
LENGTH: 14  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Sequence is synthesized  
US-09-858-935B-132

Query Match 32.3%; Score 32; DB 11; Length 14;  
Best Local Similarity 44.4%; Pred. No. 3e+02;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 SLOWITDQY 11  
DB 5 ALOWICEKY 13

RESULT 9  
US-09-858-935B-134  
Sequence 134, Application US/09858935B  
Publication No. US20030069177A1  
GENERAL INFORMATION:  
APPLICANT: Dubague, Yves  
APPLICANT: Filvaroff, Ellen  
APPLICANT: Lowman, Henry B.  
TITLE OF INVENTION: METHOD FOR TREATING CARTILAGE DISORDERS  
FILE REFERENCE: P1794R1  
CURRENT APPLICATION NUMBER: US/09/858,935B  
CURRENT FILING DATE: 2002-07-02  
PRIOR APPLICATION NUMBER: US 60/248,985  
PRIOR FILING DATE: 2000-11-15  
PRIOR APPLICATION NUMBER: US 60/204,490  
PRIOR FILING DATE: 2000-05-16  
NUMBER OF SEQ ID NOS: 153  
SEQ ID NO 134  
LENGTH: 14

TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Sequence is synthesized  
US-09-858-935B-134

Query Match 32.3%; Score 32; DB 11; Length 14;  
Best Local Similarity 44.4%; Pred. No. 3e+02;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 SLOWITDOY 11  
DB 5 NLQWLCERY 13

RESULT 10  
US-09-962-756-1752  
Sequence 1752, Application US/09962756  
Publication No. US20030195147A1  
GENERAL INFORMATION:

APPLICANT: PILUTTA, RENKA  
APPLICANT: BRISSETTE, RENEE  
APPLICANT: BLUME, ARTHUR J.  
APPLICANT: SCHAEFER, LAUGE  
APPLICANT: BRANDT, JAKOB  
APPLICANT: GOLDSTEIN, NEIL I.  
APPLICANT: SPETZLER, JANE  
APPLICANT: OSTERGAARD, SOREN  
APPLICANT: HANSEN, PER HERTZ  
TITLE OF INVENTION: INSULIN AND IGF-1 RECEPTOR AGONISTS AND ANTAGONISTS  
FILE REFERENCE: 1878-4051US1  
CURRENT APPLICATION NUMBER: US/09/962,756  
CURRENT FILING DATE: 2001-09-24  
PRIOR APPLICATION NUMBER: 09/538,038  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: 09/146,127  
PRIOR FILING DATE: 1998-09-02  
NUMBER OF SEQ ID NOS: 2227  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1752  
LENGTH: 17  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-09-962-756-1752

Query Match 32.3%; Score 32; DB 12; Length 17;  
Best Local Similarity 40.0%; Pred. No. 3.6e+02;  
Matches 4; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 DSLQWITDOY 11  
DB 2 DELEWMLDYF 11

RESULT 11  
US-10-253-471-1752  
Sequence 1752, Application US/10253471  
Publication No. US20030236190A1  
GENERAL INFORMATION:  
APPLICANT: PILUTTA, RENKA et al.  
TITLE OF INVENTION: INSULIN AND IGF-1 RECEPTOR AGONISTS AND ANTAGONISTS  
FILE REFERENCE: 1878-4057  
CURRENT APPLICATION NUMBER: US/10/253,471  
CURRENT FILING DATE: 2002-09-24  
PRIOR APPLICATION NUMBER: 09/962,756  
PRIOR FILING DATE: 2001-09-24  
PRIOR APPLICATION NUMBER: 09/538,038  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: 09/146,127  
PRIOR FILING DATE: 1998-09-02

NUMBER OF SEQ ID NOS: 2227  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1752  
LENGTH: 17  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
US-10-253-471-1752

Query Match 32.3%; Score 32; DB 12; Length 17;  
Best Local Similarity 40.0%; Pred. No. 3.6e+02;  
Matches 4; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 DSLQWITDOY 11  
DB 2 DELEWMLDYF 11

RESULT 12  
US-09-884-696-8  
Sequence 8, Application US/0984696  
Publication No. US20030035809A1  
GENERAL INFORMATION:  
APPLICANT: GEORGE, LISLE W  
APPLICANT: ANGELOS, JOHN A  
APPLICANT: HESS, JOHN P  
TITLE OF INVENTION: MORAXELLA BOVIS CYTOTOXIN, CYTOTOXIN GENE, ANTIBODIES  
TITLE OF INVENTION: AND VACCINES FOR PREVENTION AND TREATMENT OF MORAXELLA  
FILE REFERENCE: 481.06  
CURRENT APPLICATION NUMBER: US/09/884,696  
CURRENT FILING DATE: 2001-06-19  
NUMBER OF SEQ ID NOS: 41  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 8  
LENGTH: 12  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-09-884-696-8

Query Match 31.3%; Score 31; DB 11; Length 12;  
Best Local Similarity 75.0%; Pred. No. 3.6e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 ITDQYNKE 14  
DB 1 ILSQYNKE 8

RESULT 13  
US-09-897-107-81  
Sequence 81, Application US/09897107  
Patent No. US20020137094A1  
GENERAL INFORMATION:  
APPLICANT: YAMAGISHI, AKIHIKO  
TITLE OF INVENTION: METHOD FOR IMPROVING THERMOSTABILITY OF PROTEIN, PROTEINS HAVING  
TITLE OF INVENTION: THERMOSTABILITY IMPROVED BY THE METHOD AND NUCLEIC ACIDS ENCODING  
FILE REFERENCE: 210383US0  
CURRENT APPLICATION NUMBER: US/09/897,107  
CURRENT FILING DATE: 2001-07-03  
PRIOR APPLICATION NUMBER: JP2000-201920  
PRIOR FILING DATE: 2000-07-04  
PRIOR APPLICATION NUMBER: JP2001-164332  
PRIOR FILING DATE: 2001-05-31  
NUMBER OF SEQ ID NOS: 104  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 81  
LENGTH: 13  
TYPE: PRT  
ORGANISM: Bos Taurus  
US-09-897-107-81

Query Match 31.3%; Score 31; DB 10; Length 13;  
 Best Local Similarity 46.2%; Pred. No. 3.9e+02;  
 Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 6 WITDQYNKESDDK 18  
 Db 1 WMIPPEAKESNDK 13

QY 4 LOWITDOY 11  
 Db 6 LOWLCEXY 13

Search completed: January 21, 2004, 12:25:09  
 Job time : 32 secs

# RESULT 14

US-09-858-935B-29  
 ; Sequence 29; Application US/09858935B  
 ; Publication No. US2003069177A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dubaqui, Yves  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Lowman, Henry B.  
 ; TITLE OF INVENTION: METHOD FOR TREATING CARTILAGE DISORDERS  
 ; FILE REFERENCE: P1794R1  
 ; CURRENT APPLICATION NUMBER: US/09/858,935B  
 ; CURRENT FILING DATE: 2002-07-02  
 ; PRIOR APPLICATION NUMBER: US 60/248,985  
 ; PRIOR FILING DATE: 2000-11-15  
 ; PRIOR APPLICATION NUMBER: US 60/204,490  
 ; PRIOR FILING DATE: 2000-05-16  
 ; NUMBER OF SEQ ID NOS: 153  
 ; SEQ ID NO 29  
 ; LENGTH: 14  
 ; TYPE: PRT  
 ; ORGANISM: Artificial sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Sequence is synthesized  
 US-09-858-935B-29

Query Match 31.3%; Score 31; DB 11; Length 14;  
 Best Local Similarity 45.5%; Pred. No. 4.2e+02;  
 Matches 5; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 KDSLOWITDOY 11  
 Db 3 KGPLQWLCEXY 13

RESULT 15  
 US-09-858-935B-31  
 ; Sequence 31; Application US/09858935B  
 ; Publication No. US2003069177A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dubaqui, Yves  
 ; APPLICANT: Filvaroff, Ellen  
 ; APPLICANT: Lowman, Henry B.  
 ; TITLE OF INVENTION: METHOD FOR TREATING CARTILAGE DISORDERS  
 ; FILE REFERENCE: P1794R1  
 ; CURRENT APPLICATION NUMBER: US/09/858,935B  
 ; CURRENT FILING DATE: 2002-07-02  
 ; PRIOR APPLICATION NUMBER: US 60/248,985  
 ; PRIOR FILING DATE: 2000-11-15  
 ; PRIOR APPLICATION NUMBER: US 60/204,490  
 ; PRIOR FILING DATE: 2000-05-16  
 ; NUMBER OF SEQ ID NOS: 153  
 ; SEQ ID NO 31  
 ; LENGTH: 14  
 ; TYPE: PRT  
 ; ORGANISM: Artificial sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Sequence is synthesized  
 US-09-858-935B-31

Query Match 31.3%; Score 31; DB 11; Length 14;  
 Best Local Similarity 50.0%; Pred. No. 4.2e+02;  
 Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: January 21, 2004, 12:17:46 ; Search time 22 Seconds  
(without alignments)  
34.618 Million cell updates/sec

Title: US-09-941-314-7  
Perfect score: 99  
Sequence: 1 KDSLOWITTDQYKESDDK 18

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 142543

Minimum DB seq length: 0  
Maximum DB seq length: 18

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PTCUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34.3	34.3	14	3	US-09-052-888-85
2	33.3	33.3	14	4	US-08-695-692B-12
3	32.3	32.3	14	4	US-09-337-227C-4
4	32.3	32.3	14	4	US-09-337-227C-6
5	31.3	31.3	13	3	US-09-155-941-19
6	31.3	31.3	14	3	US-09-052-888-84
7	31.3	31.3	14	3	US-09-052-888-86
8	31.3	31.3	14	3	US-09-052-888-89
9	31.3	31.3	14	3	US-09-052-888-90
10	31.3	31.3	14	3	US-09-052-888-96
11	31.3	31.3	14	4	US-09-337-227C-5
12	31.3	31.3	14	4	US-09-337-227C-37
13	31.3	31.3	14	4	US-09-337-227C-38
14	31.3	31.3	15	3	US-08-825-852-15
15	31.3	31.3	15	3	US-09-052-888-15
16	31.3	31.3	15	4	US-09-337-227C-41
17	31.3	31.3	17	3	US-09-052-888-100
18	31.3	31.3	18	3	US-09-052-888-73
19	31.3	31.3	18	3	US-09-052-888-88
20	31.3	31.3	18	3	US-09-052-888-98
21	31.3	31.3	18	4	US-09-337-227C-2
22	30.3	30.3	14	1	US-08-311-611A-216
23	30.3	30.3	14	1	US-08-372-783-216
24	30.3	30.3	14	1	US-08-372-105-216
25	30.3	30.3	14	1	US-08-306-473A-216
26	30.3	30.3	14	2	US-08-621-803-133
27	30.3	30.3	14	2	US-08-485-445A-216

28	30	30.3	14	2	US-08-621-259A-104	Sequence 104, App
29	30	30.3	14	3	US-09-119-263-216	Sequence 216, App
30	30	30.3	14	3	US-08-657-162-216	Sequence 216, App
31	30	30.3	14	3	US-09-224-480-216	Sequence 216, App
32	30	30.3	14	3	US-09-217-352-133	Sequence 133, App
33	30	30.3	14	5	PCT-US95-00498-216	Sequence 216, App
34	30	30.3	14	5	PCT-US95-00656-216	Sequence 216, App
35	30	30.3	14	5	PCT-US95-09262-104	Sequence 104, App
36	30	30.3	16	4	US-09-418-780A-15	Sequence 15, App
37	30	30.3	17	4	US-09-459-958-4	Sequence 4, App
38	29	29.3	10	4	US-09-337-227C-43	Sequence 43, App
39	29	29.3	11	4	US-09-337-227C-44	Sequence 44, App
40	29	29.3	11	4	US-09-337-227C-45	Sequence 45, App
41	29	29.3	12	4	US-09-337-227C-35	Sequence 35, App
42	29	29.3	14	4	US-09-337-227C-34	Sequence 34, App
43	29	29.3	14	4	US-09-337-227C-36	Sequence 36, App
44	29	29.3	16	4	US-09-418-780A-7	Sequence 7, App
45	28	28.3	11	1	US-08-616-855-10	Sequence 10, App

## ALIGNMENTS

RESULT 1  
US-09-052-888-85  
Sequence 85, Application US/09052888  
Patent No. 6251865  
GENERAL INFORMATION:  
APPLICANT: Clark, Rosa G1  
APPLICANT: Lowman, Henry B.  
TITLE OF INVENTION: Insulin-like Growth Factor Agonist Molecules  
NUMBER OF SEQUENCES: 109  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/052,888  
FILING DATE: 31-Mar-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Haasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: P1071P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-1896  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 85:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-09-052-888-85  
Query Match 34.3%; Score 34; DB 3; Length 14;  
Best Local Similarity 45.5%; Pred. No. 34;  
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
QY 1 KDSLOWITTDQY 11  
DB 3 KGPLQWLCERY 13  
RESULT 2

US-08-695-6928-12  
; Sequence 12, Application US/086956928  
; Patent No. 6514498  
; GENERAL INFORMATION:  
; APPLICANT: Per Antonsson, Per Bjork, Mikael Dohltten,  
; APPLICANT: Johan Hansson, Terje Kalland, Lars  
; APPLICANT: Abrahamson and Goran Forsberg  
; TITLE OF INVENTION: MODIFIED/CHIMERIC SUPERANTIGENS  
; TITLE OF INVENTION: AND THEIR USE  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pravel, Hewlett, Kimball & Krieger  
; STREET: 1177 West Loop South, 10th Floor  
; CITY: Houston  
; STATE: TX  
; COUNTRY: USA  
; ZIP: 77027-9095  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/695,6928  
; FILING DATE: August 12, 1996  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 9601245-5  
; FILING DATE: March 29, 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Krieger, Paul E.  
; REGISTRATION NUMBER: 25,886  
; REFERENCE/DOCKET NUMBER: 41986/1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 713-850-0909  
; TELEFAX: 713-850-0165  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 14 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-695-6928-12

Query Match 33.3%; Score 33; DB 4; Length 14;  
Best Local Similarity 85.7%; Pred. No. 48;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
DB 4 NKESDDQ 10

RESULT 3  
US-09-337-227C-4  
; Sequence 4, Application US/09337227C  
; Patent No. 6420518  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Yvonne May-Yee  
; APPLICANT: Clark, Ross G.  
; APPLICANT: Cochran, Andrea G.  
; APPLICANT: Lowman, Henry B.  
; APPLICANT: Robinson, Iain C.A.F.  
; APPLICANT: Skelton, Nicholas J.  
; TITLE OF INVENTION: INSULIN-LIKE GROWTH FACTOR AGONIST MOLECULES  
; FILE REFERENCE: P1071P2.rev  
; CURRENT APPLICATION NUMBER: US/09/337,227C  
; CURRENT FILING DATE: 1999-06-22  
; PRIOR APPLICATION NUMBER: US 09/052,888  
; PRIOR FILING DATE: 1998-03-31  
; PRIOR APPLICATION NUMBER: US 08/825,852  
; PRIOR FILING DATE: 1997-04-04  
; NUMBER OF SEQ ID NOS: 51

; SEQ ID NO 4  
; LENGTH: 14  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Sequence is synthesized  
; Patent No. 6420518  
US-09-337-227C-4

Query Match 32.3%; Score 32; DB 4; Length 14;  
Best Local Similarity 44.4%; Pred. No. 68;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
DB 5 SLOWTCCKY 13

RESULT 4  
US-09-337-227C-6  
; Sequence 6, Application US/09337227C  
; Patent No. 6420518  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Yvonne May-Yee  
; APPLICANT: Clark, Ross G.  
; APPLICANT: Cochran, Andrea G.  
; APPLICANT: Lowman, Henry B.  
; APPLICANT: Robinson, Iain C.A.F.  
; APPLICANT: Skelton, Nicholas J.  
; TITLE OF INVENTION: INSULIN-LIKE GROWTH FACTOR AGONIST MOLECULES  
; FILE REFERENCE: P1071P2.rev  
; CURRENT APPLICATION NUMBER: US/09/337,227C  
; CURRENT FILING DATE: 1999-06-22  
; PRIOR APPLICATION NUMBER: US 09/052,888  
; PRIOR FILING DATE: 1998-03-31  
; PRIOR APPLICATION NUMBER: US 08/825,852  
; PRIOR FILING DATE: 1997-04-04  
; NUMBER OF SEQ ID NOS: 51  
; SEQ ID NO 6  
; LENGTH: 14  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Sequence is synthesized  
; Patent No. 6420518  
US-09-337-227C-6

Query Match 32.3%; Score 32; DB 4; Length 14;  
Best Local Similarity 44.4%; Pred. No. 68;  
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;  
DB 3 SLOWITDOY 11

RESULT 5  
US-09-155-941-19  
; Sequence 19, Application US/09155941  
; Patent No. 6280975  
; GENERAL INFORMATION:  
; APPLICANT: EHLERS, Marc  
; APPLICANT: GROTZINGER, Joachim  
; TITLE OF INVENTION: IL-6 MUTAIN  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 624 Ninth Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20001

COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/155,941  
FILING DATE: 06-Apr-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP96/01506  
FILING DATE: 09-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: YUN, Allen C.  
REGISTRATION NUMBER: 37,971  
REFERENCE/DOCKET NUMBER: EILERS=1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 13 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 19:  
US-09-155-941-19

Query Match 31.3%; Score 31; DB 3; Length 13;  
Best Local Similarity 60.0%; Pred. No. 89;  
Matches 6; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 9 DOYNKESDDK 18  
| | | | |  
| | | | |  
Db 3 DYNKCEDSK 12

RESULT 6  
US-09-888-84  
Sequence 84, Application US/09052888  
Patent No. 6251865  
GENERAL INFORMATION:  
APPLICANT: Clark, Ross G1  
APPLICANT: Lowman, Henry B.  
APPLICANT: Robinson, Iain C.A.F.  
TITLE OF INVENTION: Insulin-like Growth Factor Agonist Molecules  
NUMBER OF SEQUENCES: 109  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/052,888  
FILING DATE: 31-Mar-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Haarak, Janet E. 28,616  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: P1071P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/952-9881  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 84:  
SEQUENCE CHARACTERISTICS:

LENGTH: 14 amino acids  
TYPE: Amino Acid  
TOPOLOGY: linear  
US-09-052-888-84

Query Match 31.3%; Score 31; DB 3; Length 14;  
Best Local Similarity 45.5%; Pred. No. 97;  
Matches 5; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 KSLQWITDOY 11  
| | | | |  
| | | | |  
Db 3 KGPLQWLCELY 13

RESULT 7  
US-09-052-888-86  
Sequence 86, Application US/09052888  
Patent No. 6251865  
GENERAL INFORMATION:  
APPLICANT: Clark, Ross G1  
APPLICANT: Lowman, Henry B.  
APPLICANT: Robinson, Iain C.A.F.  
TITLE OF INVENTION: Insulin-like Growth Factor Agonist Molecules  
NUMBER OF SEQUENCES: 109  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/052,888  
FILING DATE: 31-Mar-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Haarak, Janet E. 28,616  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: P1071P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/952-9881  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 86:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: Amino Acid  
TOPOLOGY: linear  
US-09-052-888-86

Query Match 31.3%; Score 31; DB 3; Length 14;  
Best Local Similarity 50.0%; Pred. No. 97;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 LQWITDOY 11  
| | | | |  
| | | | |  
Db 6 LQWLCELY 13

RESULT 8  
US-09-052-888-89  
Sequence 89, Application US/09052888  
Patent No. 6251865  
GENERAL INFORMATION:  
APPLICANT: Clark, Ross G1  
APPLICANT: Lowman, Henry B.  
APPLICANT: Robinson, Iain C.A.F.  
TITLE OF INVENTION: Insulin-like Growth Factor Agonist Molecules  
NUMBER OF SEQUENCES: 109

CORRESPONDENCE ADDRESS:  
ADDRESS: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Winpatin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/052,888  
FILING DATE: 31-Mar-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: P1071P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-1896  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 89:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-09-052-888-89

Query Match 31.3%; Score 31; DB 3; Length 14;  
Best Local Similarity 50.0%; Pred. No. 97;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 LOWITDOY 11  
|||::||  
Db 6 LOWICERY 13

RESULT 9  
US-09-052-888-90  
Sequence 90, Application US/09052888  
Patent No. 6251865  
GENERAL INFORMATION:  
APPLICANT: Clark, Ross G1  
APPLICANT: Lowman, Henry B.  
APPLICANT: Robinson, Iain C.A.F.  
TITLE OF INVENTION: Insulin-like Growth Factor Agonist Molecules  
NUMBER OF SEQUENCES: 109  
CORRESPONDENCE ADDRESS:  
ADDRESS: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Winpatin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/052,888  
FILING DATE: 31-Mar-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: P1071P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-1896  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 90:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-09-052-888-90

SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-09-052-888-90

Query Match 31.3%; Score 31; DB 3; Length 14;  
Best Local Similarity 50.0%; Pred. No. 97;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 LOWITDOY 11  
|||::||  
Db 6 LOWICERY 13

RESULT 10  
US-09-052-888-96  
Sequence 96, Application US/09052888  
Patent No. 6251865  
GENERAL INFORMATION:  
APPLICANT: Clark, Ross G1  
APPLICANT: Lowman, Henry B.  
APPLICANT: Robinson, Iain C.A.F.  
TITLE OF INVENTION: Insulin-like Growth Factor Agonist Molecules  
NUMBER OF SEQUENCES: 109  
CORRESPONDENCE ADDRESS:  
ADDRESS: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Winpatin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/052,888  
FILING DATE: 31-Mar-1998  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Hasak, Janet E.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: P1071P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-1896  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 96:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-09-052-888-96

Query Match 31.3%; Score 31; DB 3; Length 14;  
Best Local Similarity 50.0%; Pred. No. 97;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 LOWITDOY 11  
|||::||  
Db 6 LOWICERY 13

RESULT 11  
US-09-337-227C-5  
Sequence 5, Application US/09337227C  
Patent No. 6420518  
GENERAL INFORMATION:  
APPLICANT: Chen, Yvonne May-Yee  
APPLICANT: Clark, Ross G.  
APPLICANT: Cochran, Andrea G.  
APPLICANT: Lowman, Henry B.

APPLICANT: Robinson, Iain C.A.F.  
APPLICANT: Skelton, Nicholas J.  
TITLE OF INVENTION: INSULIN-LIKE GROWTH FACTOR AGONIST MOLECULES  
FILE REFERENCE: P1071P2.rev  
CURRENT APPLICATION NUMBER: US/09/337,227C  
CURRENT FILING DATE: 1999-06-22  
PRIOR APPLICATION NUMBER: US 09/052,888  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: US 08/825,852  
PRIOR FILING DATE: 1997-04-04  
NUMBER OF SEQ ID NOS: 51  
SEQ ID NO 5  
LENGTH: 14  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Sequence is synthesized  
Patent No. 6420518  
US-09-337-227C-5

Query Match 31.3%; Score 31; DB 4; Length 14;  
Best Local Similarity 50.0%; Pred. No. 97;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 LOWITDOY 11  
|||: :||  
Db 6 LOWICERY 13

RESULT 12  
US-09-337-227C-37  
Sequence 37, Application US/09337227C  
Patent No. 6420518  
GENERAL INFORMATION:  
APPLICANT: Chen, Yvonne May-Yee  
APPLICANT: Clark, Ross G.  
APPLICANT: Cochran, Andrea G.  
APPLICANT: Lowman, Henry B.  
APPLICANT: Robinson, Iain C.A.F.  
APPLICANT: Skelton, Nicholas J.  
TITLE OF INVENTION: INSULIN-LIKE GROWTH FACTOR AGONIST MOLECULES  
FILE REFERENCE: P1071P2.rev  
CURRENT APPLICATION NUMBER: US/09/337,227C  
CURRENT FILING DATE: 1999-06-22  
PRIOR APPLICATION NUMBER: US 09/052,888  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: US 08/825,852  
PRIOR FILING DATE: 1997-04-04  
NUMBER OF SEQ ID NOS: 51  
SEQ ID NO 37  
LENGTH: 14  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Sequence is synthesized  
Patent No. 6420518  
US-09-337-227C-37

Query Match 31.3%; Score 31; DB 4; Length 14;  
Best Local Similarity 50.0%; Pred. No. 97;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 LOWITDOY 11  
|||: :||  
Db 6 LOWICERY 13

RESULT 13  
US-09-337-227C-38  
Sequence 38, Application US/09337227C  
Patent No. 6420518  
GENERAL INFORMATION:  
APPLICANT: Chen, Yvonne May-Yee

APPLICANT: Clark, Ross G.  
APPLICANT: Cochran, Andrea G.  
APPLICANT: Lowman, Henry B.  
APPLICANT: Robinson, Iain C.A.F.  
APPLICANT: Skelton, Nicholas J.  
TITLE OF INVENTION: INSULIN-LIKE GROWTH FACTOR AGONIST MOLECULES  
FILE REFERENCE: P1071P2.rev  
CURRENT APPLICATION NUMBER: US/09/337,227C  
CURRENT FILING DATE: 1999-06-22  
PRIOR APPLICATION NUMBER: US 09/052,888  
PRIOR FILING DATE: 1998-03-31  
PRIOR APPLICATION NUMBER: US 08/825,852  
PRIOR FILING DATE: 1997-04-04  
NUMBER OF SEQ ID NOS: 51  
SEQ ID NO 38  
LENGTH: 14  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Sequence is synthesized  
Patent No. 6420518  
US-09-337-227C-38

Query Match 31.3%; Score 31; DB 4; Length 14;  
Best Local Similarity 50.0%; Pred. No. 97;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 LOWITDOY 11  
|||: :||  
Db 6 LOWICERY 13

RESULT 14  
US-08-825-852-15  
Sequence 15, Application US/08825852  
Patent No. 6121416  
GENERAL INFORMATION:  
APPLICANT: Clark, Ross G.  
APPLICANT: Lowman, Henry B.  
APPLICANT: Robinson, Iain C.A.F.  
TITLE OF INVENTION: Insulin-like Growth Factor Agonist Molecules  
NUMBER OF SEQUENCES: 79  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 1 DNA Way  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/825,852  
FILING DATE: 04-Apr-1997  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Haak, Janet B.  
REGISTRATION NUMBER: 28,616  
REFERENCE/DOCKET NUMBER: P1071  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-1896  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
US-08-825-852-15

Query Match 31.3%; Score 31; DB 3; Length 15;

Best Local Similarity 50.0%; Pred. No. 1e+02;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 LOWITDOY 11  
|||::|  
Db 6 LOWLCEKY 13

## RESULT 15

US-09-052-888-15  
; Sequence 15, Application US/09052888  
; Patent No. 6231865  
; GENERAL INFORMATION:  
; APPLICANT: Clark, Rose G1  
; APPLICANT: Lowman, Henry B.  
; APPLICANT: Robinson, Iain C.A.F.  
; TITLE OF INVENTION: Insulin-like Growth Factor Agonist Molecules  
; NUMBER OF SEQUENCES: 109  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Winpatin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/052,888  
; FILING DATE: 31-Mar-1998  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Haasak, Janet E.  
; REGISTRATION NUMBER: 28,616  
; REFERENCE/DOCKET NUMBER: P1071P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 650/225-1896  
; TELEFAX: 650/952-9881  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15 amino acids  
; TYPE: Amino Acid  
; TOPOLOGY: Linear  
US-09-052-888-15

Query Match 31.3%; Score 31; DB 3; Length 15;  
Best Local Similarity 50.0%; Pred. No. 1e+02;  
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 4 LOWITDOY 11  
|||::|  
Db 6 LOWLCEKY 13

Search completed: January 21, 2004, 12:20:51  
Job time : 22 secs



CC quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.  
CC The polynucleotide encoding (1) is useful to detect and to localise the  
CC expression of a Zcys8 gene in a biological sample and Zcys8  
CC oligonucleotide probes are useful for in vivo diagnosis. The  
CC polynucleotide encoding (1) is useful in determining whether a subject's  
CC chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene  
CC copy number changes, insertions, deletions, restriction site changes and  
CC rearrangements and genetic alterations that inactivate the Zcys8 gene.  
CC This sequence represents an antigenic peptide of human cystatin-8 (Zcys8)  
XX  
SQ Sequence 35 AA;  
Query Match 100.0%; Score 187; DB 5; Length 35;  
Best Local Similarity 100.0%; Pred. No. 4.5e-19;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 KDSLQWITDQYNKESDDKHFRIFRVLKVRQVTD 35  
Db 1 KDSLQWITDQYNKESDDKHFRIFRVLKVRQVTD 35  
RESULT 2  
AAU79853 standard; protein; 115 AA.  
XX  
AC AAU79853;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Human cystatin-8 (Zcys8) antigenic fragment #1.  
XX  
KM Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;  
KM spermatogenesis; seminal fluid viscosity; cryopreserved sperm;  
KM sperm motility; fertilisation; antigenic fragment.  
XX  
OS Homo sapiens.  
XX  
PN WO200220567-A2.  
XX  
PD 14-MAR-2002.  
XX  
PF 29-AUG-2001; 2001WO-US0266868.  
XX  
PR 01-SEP-2000; 2000US-0230230P.  
XX  
PA (ZYMO) ZYMOGENETICS INC.  
XX  
PI Holloway JL, Gao Z, Bishop PD;  
XX  
DR WPI; 2002-383044/41.  
XX  
PT Novel isolated mammalian cystatin-8 polypeptide useful for promoting  
PT spermatogenesis, and inhibiting cancer procoagulant protein which leads  
PT to inhibition of thrombotic events associated with cancer.  
XX  
PS Claim 2; Page 94; 100pp; English.  
XX  
CC The invention describes an isolated mammalian cystatin-8 (Zcys8)  
CC polypeptide (1). (1) is useful for: inhibiting cancer procoagulant  
CC protein in an individual and thus inhibiting the thrombotic events  
CC associated with cancer; promoting spermatogenesis; modulating seminal  
CC fluid viscosity; enhancing viability of cryopreserved sperm; sperm  
CC motility and fertilisation; and as antigenic peptides to generate  
CC antibodies. Zcys8 is useful as research reagent for characterising sites  
CC of interaction between Zcys8 and its receptor. Zcys8 is useful in  
CC enhancing fertilisation during assisted reproduction in humans and in  
CC animals. Anti-(1) antibodies are useful to screen biological samples like  
CC blood, urine, saliva, tissue biopsy and autopsy material in vitro for the  
CC presence of Zcys8. The antibodies are also useful to isolate large  
CC quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.  
CC The polynucleotide encoding (1) is useful to detect and to localise the  
CC expression of a Zcys8 gene in a biological sample and Zcys8  
CC oligonucleotide probes are useful for in vivo diagnosis. The

CC polynucleotide encoding (1) is useful in determining whether a subject's  
CC chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene  
CC copy number changes, insertions, deletions, restriction site changes and  
CC rearrangements and genetic alterations that inactivate the Zcys8 gene.  
CC This sequence represents an antigenic fragment of human cystatin-8  
CC (Zcys8)  
XX  
SQ Sequence 115 AA;  
Query Match 100.0%; Score 187; DB 5; Length 115;  
Best Local Similarity 100.0%; Pred. No. 1.8e-18;  
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 KDSLQWITDQYNKESDDKHFRIFRVLKVRQVTD 35  
Db 22 KDSLQWITDQYNKESDDKHFRIFRVLKVRQVTD 56  
RESULT 3  
AAU79854 standard; protein; 117 AA.  
XX  
AC AAU79854;  
XX  
DT 15-JUL-2002 (first entry)  
XX  
DE Human cystatin-8 (Zcys8) antigenic fragment #2.  
XX  
KM Cystatin-8; Zcys8; cancer; procoagulant protein; thrombosis;  
KM spermatogenesis; seminal fluid viscosity; cryopreserved sperm;  
KM sperm motility; fertilisation; antigenic fragment.  
XX  
OS Homo sapiens.  
XX  
PN WO200220567-A2.  
XX  
PD 14-MAR-2002.  
XX  
PF 29-AUG-2001; 2001WO-US0266868.  
XX  
PR 01-SEP-2000; 2000US-0230230P.  
XX  
PA (ZYMO) ZYMOGENETICS INC.  
XX  
PI Holloway JL, Gao Z, Bishop PD;  
XX  
DR WPI; 2002-383044/41.  
XX  
PT Novel isolated mammalian cystatin-8 polypeptide useful for promoting  
PT spermatogenesis, and inhibiting cancer procoagulant protein which leads  
PT to inhibition of thrombotic events associated with cancer.  
XX  
PS Claim 2; Page 94-95; 100pp; English.  
XX  
CC The invention describes an isolated mammalian cystatin-8 (Zcys8)  
CC polypeptide (1). (1) is useful for: inhibiting cancer procoagulant  
CC protein in an individual and thus inhibiting the thrombotic events  
CC associated with cancer; promoting spermatogenesis; modulating seminal  
CC fluid viscosity; enhancing viability of cryopreserved sperm; sperm  
CC motility and fertilisation; and as antigenic peptides to generate  
CC antibodies. Zcys8 is useful as research reagent for characterising sites  
CC of interaction between Zcys8 and its receptor. Zcys8 is useful in  
CC enhancing fertilisation during assisted reproduction in humans and in  
CC animals. Anti-(1) antibodies are useful to screen biological samples like  
CC blood, urine, saliva, tissue biopsy and autopsy material in vitro for the  
CC presence of Zcys8. The antibodies are also useful to isolate large  
CC quantities of Zcys8 protein and DNA sequences that encode Zcys8 genes.  
CC The polynucleotide encoding (1) is useful to detect and to localise the  
CC expression of a Zcys8 gene in a biological sample and Zcys8  
CC oligonucleotide probes are useful for in vivo diagnosis. The  
CC polynucleotide encoding (1) is useful in determining whether a subject's  
CC chromosomes contain a mutation in the Zcys8 gene like aneuploidy, gene  
CC copy number changes, insertions, deletions, restriction site changes and

[illegible]

```

XX      Sequence 137 AA;
SQ
Query Match      100.0%; Score 187, DB 5, Length 137,
Best Local Similarity 100.0%; Pred. No. 2, 2e-18;
Matches 35, Conservative 0, Mismatches 0; Indels 0; Gaps 0;

QY      1 KDSLQWITPDQYNKESDDKXFRIRFVRLKVQQRVD 35
Db      44 KDSLQWITPDQYNKESDDKXFRIRFVRLKVQQRVD 78

RESULT 5
AAM15096
ID      AAM15096 standard; protein, 50 AA.
XX
AC      AAM15096;
XX
DT      12-OCT-2001 (first entry)
XX
DE      Peptide #1530 encoded by probe for measuring cervical gene expression.
XX
KW      Probe; human; microarray; gene expression; cervical epithelial cell;
KM      cervical cancer.
XX
OS      Homo sapiens.
XX
PN      WO200157278-A2.
PD
PF      09-AUG-2001.
XX
PR      30-JAN-2001; 2001WO-US000670.
XX
PR      04-FEB-2000; 2000US-0180312P.
XX      26-MAY-2000; 2000US-0207456P.
PR      30-JUN-2000; 2000US-00608408.
PR      03-AUG-2000; 2000US-00632366.
PR      21-SEP-2000; 2000US-0234687P.
PR      27-SEP-2000; 2000US-0236359P.
PR      04-OCT-2000; 2000GB-00024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR      WPI; 2001-488901/53.
XX
PT      Human genome-derived single exon nucleic acid probes useful for analyzing
PT      gene expression in human cervical epithelial cells.
XX
PS      Claim 27; SEQ ID NO 19922; 487pp; English.
XX
CC      The present invention relates to human single exon nucleic acid probes
CC      (SENP: see AAI10068-AA128459). The present sequence is a peptide encoded
CC      by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC      can be used to produce a single exon microarray, which can be used for
CC      measuring human gene expression in a sample derived from human cervical
CC      epithelial cells. By measuring gene expression, the probes are therefore
CC      useful in grading and/or staging of diseases of the cervix, notably
CC      cervical cancer. Note: The sequence data for this patent did not form
CC      part of the printed specification, but was obtained in electronic format
CC      directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ      Sequence 50 AA;

Query Match      92.0%; Score 172, DB 4, Length 50;
Best Local Similarity 100.0%; Pred. No. 8, 8e-17;
Matches 32, Conservative 0, Mismatches 0; Indels 0; Gaps 0;

QY      1 KDSLQWITPDQYNKESDDKXFRIRFVRLKVQQRQ 32
Db      19 KDSLQWITPDQYNKESDDKXFRIRFVRLKVQQRQ 50

```

```

RESULT 6
ABBB34086
ID ABBB34086 standard; peptide; 50 AA.
XX
AC ABB34086;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #1592 encoded by human foetal liver single exon probe.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS Homo sapiens.
PN MO200157277-A2.
XX
FD 09-AUG-2001.
XX
PR 30-JAN-2001; 2001WO-US000669.
XX
PE 04-FEB-2000; 2000US-0180312P.
XX
PR 26-MAY-2000; 2000US-0207456P.
XX
PR 30-JUN-2000; 2000US-00608408.
XX
PR 03-AUG-2000; 2000US-00632366.
XX
PR 21-SEP-2000; 2000US-0234687P.
XX
PR 27-SEP-2000; 2000US-0236359P.
XX
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PS gene expression in human fetal liver.
XX
PS Claim 27; SEQ ID NO 26721; 639pp + Sequence Listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human fetal liver. The
CC present sequence is a peptide encoded by a single exon nucleic acid probe
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 50 AA;
XX
Query Match          92.0%; Score 172; DB 4; Length 50;
Best Local Similarity 100.0%; Pred. No. 8,8e-17;
Matches   32; Conservative    0; Mismatches    0; Indels    0; Gaps    0
QY      1 KDSLQWITDYKNKESDDKKHFRIFFVLKVQRQ 32
Db      19 KDSLQWITDYKNKESDDKKHFRIFFVLKVQRQ 50

```

OS	Homo sapiens.
XX	
PN	WO200157272-A2.
XX	
PD	09-AUG-2001.
PF	
XX	
PR	30-JAN-2001; 2001WO-US000663.
XX	
PR	04-FEB-2000; 2000US-0180312P.
PR	26-MAY-2000; 2000US-0207456P.
PR	30-JUN-2000; 2000US-00608408.
PR	03-AUG-2000; 2000US-00632366.
PR	21-SEP-2000; 2000US-0234687P.
PR	27-SEP-2000; 2000US-0236359P.
PR	04-OCT-2000; 2000GB-00024263.
PA	(MOLE-) MOLECULAR DYNAMICS INC.
XX	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	
DR	WPI; 2001-488897/53.
XX	
PT	Human genome-derived single exon nucleic acid probes useful for analyzing
XX	gene expression in human placenta.
PS	Claim 27; SEQ ID NO 27814; 654bp; English.
XX	
CC	The present invention relates to single exon nucleic acid probes (SENPs;
CC	see AA13115-AA157546). The present sequence is a peptide encoded by one
CC	such probe. The probes are useful for producing a microarray for
CC	predicting, measuring and displaying gene expression in samples derived
CC	from human placenta. The probes are useful for antenatal diagnosis of
XX	human genetic disorders
SQ	Sequence 50 AA;
OY	
Db	Query Match 92.0%; Score 172; DB 4; Length 50; Best Local Similarity 100.0%; Pred. No. 8.8e-17; Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0.  1 KDSLQWITDQYNKESDDKYHFRIRVLKVQRQ 32     19 KDSLQWITDQYNKESDDKYHFRIRVLKVQRQ 50
RESULT 8	
ABB32389	
ID	ABB32389 standard; peptide; 50 AA.
XX	
AC	ABB32389;
XX	
DT	01-FEB-2002 (first entry)
XX	
DE	Peptide #5040 encoded by breast cell single exon nucleic acid probe.
KM	Human; microarray; single exon probe; gene expression; breast; disease; cancer.
XX	
OS	Homo sapiens.
XX	
PN	WO200157271-A2.
XX	
PD	09-AUG-2001.
XX	
PF	30-JAN-2001; 2001WO-US000662.
XX	
PR	04-FEB-2000; 2000US-0180312P.
PR	26-MAY-2000; 2000US-0207456P.
PR	30-JUN-2000; 2000US-00608408.
PR	03-AUG-2000; 2000US-00632366.
PR	21-SEP-2000; 2000US-0234687P.
PR	27-SEP-2000; 2000US-0236359P.
PR	04-OCT-2000; 2000GB-00024263.

```

XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX New spatially-addressable set of single exon nucleic acid probes, useful
PT for measuring gene expression in sample derived from human breast,
PT comprises number of single exon nucleic acid probes.
XX Claim 27; SEQ ID NO 15357; 327bp + Sequence Listing; English.
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting the
CC probes with a collection of detectably labeled nucleic acids derived
CC from mRNA of human breast, and then measuring the label bound to each
CC probe of the microarray. The probes are useful for verifying the
CC expression of regions of genomic DNA predicted to encode proteins. They
CC are useful for gene discovery, and for determining predisposition and/or
CC prognosing breast disease. Gene expression analysis is useful for
CC assessing the toxicity of chemical agents on cells. The microarray of
CC this invention presents a far greater diversity of probes for measuring
CC gene expression, with far less bias than expressed sequence tag
CC microarrays. The method is suitable for rapid production of functional
CC information from genomic sequence. The present sequence is a peptide
CC encoded by a single exon nucleic acid probe of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 50 AA;
SQ
Query Match 92.0%; Score 172; DB 4; Length 50;
Best Local Similarity 100.0%; Pred. No. 8.8e-17;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KDSIQWITDOYNKESDDKHFRIFRVLKVRQ 32
Db 19 KDSIQWITDOYNKESDDKHFRIFRVLKVRQ 50
RESULT 9
ABR28913
ID ABR28913 standard; peptide; 50 AA.
XX
XX ABR28913;
AC
XX
DT 01-FEB-2002 (first entry)
XX
XX Peptide #1564 encoded by breast cell single exon nucleic acid probe.
DE Human; microarray; single exon probe; gene expression; breast; disease;
KW cancer.
XX
XX Homo sapiens.
OS
XX
XX WO200157271-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US000662.
PF
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA

```

```

XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX New spatially-addressable set of single exon nucleic acid probes, useful
PT for measuring gene expression in sample derived from human breast,
PT comprises number of single exon nucleic acid probes.
XX Claim 27; SEQ ID NO 11881; 327bp + Sequence Listing; English.
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting the
CC probes with a collection of detectably labeled nucleic acids derived
CC from mRNA of human breast, and then measuring the label bound to each
CC probe of the microarray. The probes are useful for verifying the
CC expression of regions of genomic DNA predicted to encode proteins. They
CC are useful for gene discovery, and for determining predisposition and/or
CC prognosing breast disease. Gene expression analysis is useful for
CC assessing the toxicity of chemical agents on cells. The microarray of
CC this invention presents a far greater diversity of probes for measuring
CC gene expression, with far less bias than expressed sequence tag
CC microarrays. The method is suitable for rapid production of functional
CC information from genomic sequence. The present sequence is a peptide
CC encoded by a single exon nucleic acid probe of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 50 AA;
SQ
Query Match 92.0%; Score 172; DB 4; Length 50;
Best Local Similarity 100.0%; Pred. No. 8.8e-17;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 KDSIQWITDOYNKESDDKHFRIFRVLKVRQ 32
Db 19 KDSIQWITDOYNKESDDKHFRIFRVLKVRQ 50
RESULT 10
ABR19524
ID ABR19524 standard; protein; 50 AA.
XX
XX ABR19524;
AC
XX
XX 23-JAN-2002 (first entry)
DT
XX
XX Protein #1523 encoded by probe for measuring heart cell gene expression.
DE Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease.
XX
XX Homo sapiens.
OS
XX
XX WO200157274-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US000666.
PF
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA

```

PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-48899/53.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
PT hearts.  
XX  
XX Claim 15; SEQ ID NO 21294; 530bp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart (see  
CC AB2135-ABA1135). The present sequence is a protein encoded by one such  
CC probe. The probes may be used for predicting, measuring and displaying  
CC gene expression in samples derived from the human heart via microarrays.  
CC By measuring gene expression, the probes are useful for predicting,  
CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
CC human heart and vascular system e.g. cardiovascular disease,  
CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 50 AA;  
XX  
Query Match 92.0%; Score 172; DB 4; Length 50;  
Best Local Similarity 100.0%; Pred. No. 8.8e-17;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
OY 1 KDSLOWITDQYNKESDDKYHFRIRFVLKVQRQ 32  
Db 19 KDSLOWITDQYNKESDDKYHFRIRFVLKVQRQ 50  
XX  
RESULT 11  
AAM67252  
ID AAM6752 standard; protein; 50 AA.  
XX  
AC AAM67252;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 27558.  
XX  
KM Human; bone marrow expressed exon; gene expression analysis; probe;  
KM microarray; cancer; leukaemia; lymphoma; myeloma.  
XX  
OS Homo sapiens.  
XX  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000668.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-48890/53.  
XX  
PT Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human bone marrow.  
XX  
XX Example 4; SEQ ID NO 27558; 658bp + Sequence Listing; English.  
XX

CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
CC protein encoded by one of the probes of the invention  
XX  
SQ Sequence 50 AA;  
XX  
Query Match 92.0%; Score 172; DB 4; Length 50;  
Best Local Similarity 100.0%; Pred. No. 8.8e-17;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
OY 1 KDSLOWITDQYNKESDDKYHFRIRFVLKVQRQ 32  
Db 19 KDSLOWITDQYNKESDDKYHFRIRFVLKVQRQ 50  
XX  
RESULT 12  
AAM54871  
ID AAM54871 standard; protein; 50 AA.  
XX  
AC AAM54871;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 26976.  
XX  
KM Human; brain expressed exon; gene expression analysis; probe; microarray;  
KM Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO200157275-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US000667.  
XX  
PR 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-483446/52.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
PT brains.  
XX  
PS Example 4; SEQ ID NO 26976; 650bp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is a protein encoded by one of  
CC the probes of the invention  
XX  
SQ Sequence 50 AA;  
XX  
Query Match 92.0%; Score 172; DB 4; Length 50;  
Best Local Similarity 100.0%; Pred. No. 8.8e-17;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
OY 1 KDSLOWITDQYNKESDDKYHFRIRFVLKVQRQ 32

Db 19 KDSLOWITTDQYNKESDDKXHFRIFRVLKVGORQ 50

RESULT 13  
ABG48915  
ID ABG48915 standard; peptide; 50 AA.

XX

|||||

AC ABG48915;  
XX  
XX 25-FEB-2003 (first entry)  
XX  
XX Human liver peptide, SEQ ID NO 27563.  
XX  
XX Human, liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
XX hypercholesterolaemia; coronary heart disease.  
XX  
XX Homo sapiens.  
XX  
XX MO200157273-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US000664.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
XX 26-MAY-2000; 2000US-0207456P.  
XX 30-JUN-2000; 2000US-00608408.  
XX 03-AUG-2000; 2000US-00632366.  
XX 21-SEP-2000; 2000US-0234687P.  
XX 27-SEP-2000; 2000US-0236359P.  
XX 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOL-.) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2001-488898/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
XX PT gene expression in human adult liver.  
XX  
XX PS Claim 27; SEQ ID NO 27563; 658bp; English.  
XX  
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for  
XX CC measuring human gene expression in a sample derived from human adult  
XX CC liver, comprising one of 13109 defined nucleotide sequences given in the  
XX CC specification (or complements/ fragments). The probe hybridises at high  
XX CC stringency to a nucleic acid molecule expressed in the human adult liver.  
XX CC (I) may be used for predicting, measuring and displaying gene expression  
XX CC in samples derived from human adult liver. The genes identified may be  
XX CC involved in genetic liver diseases such as cirrhosis,  
XX CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
XX CC associated with coronary heart disease. ABG47348-ABG59930 represent human  
XX CC liver single exon encoded peptides of the invention. Note: The sequence  
XX CC information for this patent does not appear in the printed specification  
XX CC but was obtained in electronic format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 50 AA;

Query Match 92.0%; Score 172; DB 4; Length 50;

Best Local Similarity 100.0%; Pred. No. 8.8e-17; Mismatches 0; Indels 0; Gaps 0;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 KDSLOWITTDQYNKESDDKXHFRIFRVLKVGORQ 32

|||||

19 KDSLOWITTDQYNKESDDKXHFRIFRVLKVGORQ 50

RESULT 14  
AAM02833  
ID AAM02833 standard; protein; 50 AA.

XX AAM02833;

XX 09-OCT-2001 (first entry)

XX Peptide #1515 encoded by probe for measuring breast gene expression.

XX Probe: human; breast disease; breast cancer; development disorder;

XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX Homo sapiens.

XX MO200157270-A2.

XX 09-AUG-2001.

XX 29-JAN-2001; 2001WO-US000661.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOL-.) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-476286/51.

XX Novel single exon nucleic acid probe used to measuring gene expression in

XX a human breast.

XX PS Claim 27; SEQ ID NO 11573; 322pp; English.

XX The present invention relates to novel single exon nucleic acid probes

XX (see AAI00010-AAI10067). The present sequence is a peptide encoded by one

XX such probe. The probes are useful for measuring human gene expression in

XX a human breast sample, where the probe hybridises at high stringency to a

XX nucleic acid expressed in the human breast. The probes are useful for

XX predicting, diagnosing, grading, staging, monitoring and prognosing

XX diseases of the human breast, particularly those diseases with polygenic

XX aetiology. The diseases include: breast cancer, disorders of development,

XX inflammatory diseases of the breast, fibrocystic changes, proliferative

XX breast disease and non-carcinoma tumours. Note: The sequence data for

XX this patent did not form part of the printed specification, but was

XX obtained in electronic format directly from WIPO at

XX ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 50 AA;

Query Match 92.0%; Score 172; DB 4; Length 50;

Best Local Similarity 100.0%; Pred. No. 8.8e-17; Mismatches 0; Indels 0; Gaps 0;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 KDSLOWITTDQYNKESDDKXHFRIFRVLKVGORQ 32

|||||

19 KDSLOWITTDQYNKESDDKXHFRIFRVLKVGORQ 50

RESULT 15  
ABG36903  
ID ABG36903 standard; peptide; 50 AA.

XX ABG36903;

XX 19-AUG-2002 (first entry)

XX Human peptide encoded by genome-derived single exon probe SEQ ID 26568.

XX Human, single exon probe; asthma; lung cancer; COPD; ILD;

KM chronic obstructive pulmonary disease; interstitial lung disease;  
 KM familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KM tuberous sclerosis; Gaucher's disease; Niemann-pick disease;  
 KM Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KM pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
 KM pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KM primary ciliary dyskinesia; pulmonary hypertension;  
 KM hyaline membrane disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200186003-A2.  
 XX  
 PD 15-NOV-2001.  
 XX  
 PF 30-JUN-2001; 2001WO-US000665.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2002-114183/15.  
 XX  
 PT Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples.  
 XX  
 PS Claim 27; SEQ ID NO 26568; 634bp; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of probes  
 CC; the novel set of probes which hybridize at high stringency to a nucleic  
 CC acid expressed in the human lung; measuring gene expression in a sample  
 CC derived from human lung; comprising (a) contacting the array with a  
 CC collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of the  
 CC array; identifying exons in a eukaryotic genome, comprising (a)  
 CC algorithmically predicting at least one exon from genomic sequences of  
 CC the eukaryote; and (b) detecting specific hybridization of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridization to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene expression  
 CC analysis and for identifying exons in a gene, particularly using human  
 CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
 CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
 CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
 CC tuberous sclerosis, Gaucher's disease, Niemann-pick disease, Hermansky-  
 CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
 CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
 CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
 CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
 CC present sequence is a peptide/protein encoded by a single exon probe of  
 CC the invention. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format

CC directly from WIFO at ftp.wifo.int/pub/published\_pcf\_sequences  
 XX  
 SQ Sequence 50 AA;  
 Query Match 92.0%; Score 172; DB 5; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 8, 8e-17;  
 Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 KDSIQWITDQYNKESDDKXHFRIFRVLKVQRQ 32  
 DB 19 KDSIQWITDQYNKESDDKXHFRIFRVLKVQRQ 50  
 Search completed: March 18, 2004, 14:14:59  
 Job time : 32.7361 secs

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

